

Long paths in first passage percolation on the complete graph I. Local PWIT dynamics

Maren Eckhoff*

Jesse Goodman[†]Remco van der Hofstad[‡]Francesca R. Nardi[‡]

December 23, 2015

Abstract

We study the random geometry of first passage percolation on the complete graph equipped with independent and identically distributed edge weights, continuing the program initiated by Bhamidi and van der Hofstad [9]. We describe our results in terms of a sequence of parameters $(s_n)_{n \geq 1}$ that quantifies the extreme-value behavior of small weights, and that describes different universality classes for first passage percolation on the complete graph. We consider both n -independent as well as n -dependent edge weights. The simplest example consists of edge weights of the form E^{s_n} , where E is an exponential random variable with mean 1.

In this paper, we investigate the case where $s_n \rightarrow \infty$, and focus on the local neighborhood of a vertex. We establish that the smallest-weight tree of a vertex locally converges to the invasion percolation cluster on the Poisson weighted infinite tree. In addition, we identify the scaling limit of the weight of the smallest-weight path between two uniform vertices.

1 Model and results

In this paper, we continue the program of studying first passage percolation on the complete graph initiated in [9]. We start by introducing first passage percolation (FPP). Given a graph $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$, let $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$ denote a collection of positive edge weights. Thinking of $Y_e^{(\mathcal{G})}$ as the cost of crossing an edge e , we can define a metric on $V(\mathcal{G})$ by setting

$$d_{\mathcal{G}, Y^{(\mathcal{G})}}(i, j) = \inf_{\pi: i \rightarrow j} \sum_{e \in \pi} Y_e^{(\mathcal{G})}, \quad (1.1)$$

where the infimum is over all paths π in \mathcal{G} that join i to j , and $Y^{(\mathcal{G})}$ represents the edge weights $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$. We will always assume that the infimum in (1.1) is attained uniquely, by some (finite) path $\pi_{i,j}$. We are interested in the situation where the edge weights $Y_e^{(\mathcal{G})}$ are *random*, so that $d_{\mathcal{G}, Y^{(\mathcal{G})}}$ is a random metric. In particular, when the graph \mathcal{G} is very large, with $|V(\mathcal{G})| = n$ say, we wish to understand the scaling behavior of the following quantities for fixed $i, j \in V(\mathcal{G})$:

*Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, United Kingdom. Email: eckhoff.maren@gmail.com

[†]Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Email: jesse.goodman@auckland.ac.nz

[‡]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. Email: rhofstad@win.tue.nl, f.r.nardi@tue.nl

- (a) The *distance* $W_n = d_{\mathcal{G}, Y^{(\mathcal{G})}}(i, j)$ – the total edge cost of the optimal path $\pi_{i,j}$;
- (b) The *hopcount* H_n – the number of edges in the optimal path $\pi_{i,j}$;
- (c) The *topological structure* – the shape of the random neighborhood of a point.

In this paper, we consider FPP on the complete graph and focus on problem (c). In the companion paper [11], we will use these results to investigate problems (a) and (b). We will often refer to results in [11], and write, e.g., [Part II, Section 5.2] to refer to [11, Section 5.2]. We also refer to [Part II, Section 1.4] for an extended discussion of the results in these two papers and their relations to the literature.

In [9], the question was raised what the *universality classes* are for this model. We bring the discussion substantially further by describing a way to distinguish several universality classes and by identifying the limiting behavior of first passage percolation in one of these classes. The cost regime introduced in (1.1) uses the information from all edges along the path and is known as the *weak disorder* regime. By contrast, in the *strong disorder* regime the cost of a path π is given by $\max_{e \in \pi} Y_e^{(\mathcal{G})}$. We establish a firm connection between the weak and strong disorder regimes in first passage percolation. Interestingly, this connection also establishes a strong relation to invasion percolation (IP) on the Poisson-weighted infinite tree (PWIT), which is the scaling limit of IP on the complete graph. This process also arises as the scaling limit of the minimal spanning tree on K_n .

Our main interest is in the case $\mathcal{G} = K_n$, the complete graph on n vertices $V(K_n) = [n] := \{1, \dots, n\}$, equipped with independent and identically distributed (i.i.d.) edge weights $(Y_e^{(K_n)})_{e \in E(K_n)}$. We write Y for a random variable with $Y \stackrel{d}{=} Y_e^{(\mathcal{G})}$, and assume that the distribution function F_Y of Y is continuous. For definiteness, we study the optimal path $\pi_{1,2}$ between vertices 1 and 2; by exchangeability, $\pi_{1,2}$ has the same law as $\pi_{u,v}$ for any other $u, v \in [n]$, $u \neq v$. In [9] and [12] this setup was studied for the case that $Y_e^{(K_n)} \stackrel{d}{=} E^s$ where E is an exponential mean 1 random variable, and $s > 0$ constant and $s = s_n > 0$ a null-sequence, respectively. We start by stating our main theorem for this situation where $s = s_n$ tends to infinity. First, we introduce some notation:

Notation. All limits in this paper are taken as n tends to infinity unless stated otherwise. A sequence of events $(\mathcal{A}_n)_n$ happens *with high probability (whp)* if $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$. For random variables $(X_n)_n, X$, we write $X_n \xrightarrow{d} X$, $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{a.s.} X$ to denote convergence in distribution, in probability and almost surely, respectively. For real-valued sequences $(a_n)_n, (b_n)_n$, we write $a_n = O(b_n)$ if the sequence $(a_n/b_n)_n$ is bounded; $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$; $a_n = \Theta(b_n)$ if the sequences $(a_n/b_n)_n$ and $(b_n/a_n)_n$ are both bounded; and $a_n \sim b_n$ if $a_n/b_n \rightarrow 1$. Similarly, for sequences $(X_n)_n, (Y_n)_n$ of random variables, we write $X_n = O_{\mathbb{P}}(Y_n)$ if the sequence $(X_n/Y_n)_n$ is tight; $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \xrightarrow{\mathbb{P}} 0$; and $X_n = \Theta_{\mathbb{P}}(Y_n)$ if the sequences $(X_n/Y_n)_n$ and $(Y_n/X_n)_n$ are both tight. Moreover, E always denotes an exponentially distributed random variable with mean 1.

1.1 First passage percolation with n -dependent edge weights

We start by investigating the case where $Y = E^{s_n}$ where $s_n \rightarrow \infty$:

Theorem 1.1 (Weight – n -dependent edge weights). *Let $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$, where $(s_n)_n$ is a positive sequence with $s_n / \log \log n \rightarrow \infty$. Let $M^{(1)}, M^{(2)}$ be i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$*

is the survival probability of a Poisson Galton–Watson branching process with mean x . Then

$$nW_n^{1/s_n} \xrightarrow{d} M^{(1)} \vee M^{(2)}. \quad (1.2)$$

When $s_n \rightarrow \infty$ the values of the random weights $E_e^{s_n}$ depend strongly on the disorder $(E_e)_{e \in E(\mathcal{G})}$, making small values increasingly more favorable and large values increasingly less favorable, thus the weak disorder problem with weights $E_e^{s_n}$ approaches the strong disorder problem. Mathematically, the elementary limit

$$\lim_{s \rightarrow \infty} (x_1^s + x_2^s)^{1/s} = x_1 \vee x_2 \quad (1.3)$$

expresses the convergence of the ℓ^s norm towards the ℓ^∞ norm and establishes a relationship between the weak disorder regime and the strong disorder regime of FPP.

We continue by discussing the result in [Theorem 1.1](#) in more detail. Any sequence $(u_n(x))_n$ for which $nF_Y(u_n(x)) \rightarrow x$ is such that, for i.i.d. random variables $(Y_i)_{i \in \mathbb{N}}$ with distribution function F_Y ,

$$\mathbb{P}\left(\min_{i \in [n]} Y_i \leq u_n(x)\right) \rightarrow 1 - e^{-x}. \quad (1.4)$$

As the distribution function F_Y is continuous, we will choose $u_n(x) = F_Y^{-1}(x/n)$, so that $nF_Y(u_n(x))$ is x . The value $u_n(1)$ is denoted by u_n . In view of (1.4), the family $(u_n(x))_{x \in (0, \infty)}$ are the *characteristic values* for $\min_{i \in [n]} Y_i$. See [13] for a detailed discussion of extreme value theory. (In the strong disorder regime, $u_n(x)$ varies heavily in x such that the phrase characteristic values can be misleading.) In the setting of [Theorem 1.1](#), $F_Y(y) = 1 - e^{-y^{1/s_n}} \approx y^{1/s_n}$ when $y = y_n$ tends to zero fast enough, so that $u_n(x)$ can be taken as $u_n(x) \approx (x/n)^{s_n}$ (where \approx indicates approximation with uncontrolled error). Then we see in (1.2) that $W_n \approx u_n(M^{(1)} \vee M^{(2)})$, which means that the random fluctuations of the weight of the smallest-weight path are of the same order of magnitude as some of the typical values for the minimal edge weight adjacent to vertices 1 and 2.

To explain the appearance of the random variables $M^{(1)}$ and $M^{(2)}$, we now informally state that the local neighborhoods of the smallest-weight tree for FPP from a single source converges to invasion percolation on the so-called Poisson-weighted infinite tree. We start by introducing these notions informally; for more details see [Section 2.1](#). In *invasion percolation* (IP) on a weighted graph, we grow the invasion percolation cluster by starting from a single vertex and sequentially adding the edge attached to the cluster with minimal edge weight. The *Poisson-weighted infinite tree* (PWIT), which serves as the large- n limit of the complete graph with i.i.d. exponential edge weights, is the infinite weighted tree for which the edge weight between a vertex and its i th child, jointly in i , has the same distribution as $E_1 + \dots + E_i$, where $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables with mean 1. The weights of edges between different vertices and their children in the tree are independent. When performing IP on the PWIT, the largest edge weight ever to be accepted has distribution $M^{(1)}$ as in [Theorem 1.1](#). When performing IP on the complete graph started from the two sources 1 and 2, the largest edge weight accepted converges in distribution to $M^{(1)} \vee M^{(2)}$. This explains how [Theorem 1.1](#) can be interpreted in terms of IP on the PWIT and on the complete graph. The next theorem shows that this relation also holds for local neighborhoods of vertices:

Theorem 1.2. *Let $Y_e^{(K_n)} \stackrel{d}{=} E_e^{s_n}$, where $(s_n)_n$ is a positive sequence with $s_n \rightarrow \infty$. For each fixed $m \in \mathbb{N}$, the topology of the FPP smallest-weight tree to the nearest m vertices, as well as the weights along its edges, converge in distribution to invasion percolation on the Poisson-weighted infinite tree.*

In [Section 2.2](#), all notions needed in [Theorem 1.2](#) are formally defined. Moreover, the evolution of two smallest-weight trees started from vertices 1 and 2 is explored in detail.

1.2 The universal picture

Our results are applicable not only to the edge-weight distribution $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$ with $s_n \rightarrow \infty$, but to a large family of distributions which we now characterize. Interestingly, this family includes edge weights whose distribution is independent of n . For fixed n , the edge weights $(Y_e^{(K_n)})_{e \in E(K_n)}$ are independent for different e . However, there is no requirement that they are independent over n , and in fact in [Section 4](#), we will produce $Y_e^{(K_n)}$ using a fixed source of randomness not depending on n . Therefore, it will be useful to describe the randomness on the edge weights $((Y_e^{(K_n)})_{e \in E(K_n)} : n \in \mathbb{N})$ uniformly across the sequence. It will be most useful to give this description in terms of exponential random variables.

Distribution function: Choose a distribution function $F_Y(y)$ having no atoms, and draw the edge weights $Y_e^{(K_n)}$ independently according to $F_Y(y)$:

$$\mathbb{P}(Y_e^{(K_n)} \leq y) = F_Y(y). \quad (1.5)$$

Parametrization by Exponential variables: Fix independent exponential mean 1 variables $(E_e^{(K_n)})_{e \in E(K_n)}$, choose a strictly increasing function $g: (0, \infty) \rightarrow (0, \infty)$, and define

$$Y_e^{(K_n)} = g(E_e^{(K_n)}). \quad (1.6)$$

The relations between these parametrizations are given by

$$F_Y(y) = 1 - e^{-g^{-1}(y)} \quad \text{and} \quad g(x) = F_Y^{-1}(1 - e^{-x}). \quad (1.7)$$

We define

$$f_n(x) = g(x/n) = F_Y^{-1}(1 - e^{-x/n}). \quad (1.8)$$

Let Y_1, \dots, Y_n be i.i.d. with $Y_i = g(E_i)$ as in (1.6). Since g is increasing,

$$\min_{i \in [n]} Y_i = g(\min_{i \in [n]} E_i) \stackrel{d}{=} g(E/n) = f_n(E). \quad (1.9)$$

Because of this convenient relation between the edge weights $Y_e^{(K_n)}$ and exponential random variables, we will express our hypotheses about the distribution of the edge weights in terms of conditions on the functions $f_n(x)$ as $n \rightarrow \infty$. In [Section 1.3](#), we use this formulation for increasing g or analogous for decreasing function, to explore the universality class in which our results hold in more detail.

Consider first the case $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$ from [Theorem 1.1](#). From (1.6), we have $g(x) = g_n(x) = x^{s_n}$, so that (1.8) yields

$$f_n(x) = (x/n)^{s_n} = f_n(1)x^{s_n}, \quad \text{for} \quad Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}. \quad (1.10)$$

Thus, (1.9)–(1.10) show that the parameter s_n measures the relative *sensitivity* of $\min_{i \in [n]} Y_i$ to fluctuations in the variable E . In general, we will have $f_n(x) \approx f_n(1)x^{s_n}$ if x is appropriately close to 1 and $s_n \approx f'_n(1)/f_n(1)$. These observations motivate the following conditions on the functions $(f_n)_n$, which we will use to relate the distributions of the edge weights $Y_e^{(K_n)}$, $n \in \mathbb{N}$, to a sequence $(s_n)_n$:

Condition 1.3 (Scaling of f_n). For every $x \geq 0$,

$$\frac{f_n(x^{1/s_n})}{f_n(1)} \rightarrow x. \quad (1.11)$$

Condition 1.4 (Density bound for small weights). There exist $\varepsilon_0 > 0$, $\delta_0 \in (0, 1]$ and $n_0 \in \mathbb{N}$ such that

$$\varepsilon_0 s_n \leq x \frac{d}{dx} \log f_n(x) \leq s_n / \varepsilon_0, \quad \text{whenever } 1 - \delta_0 \leq x \leq 1, \text{ and } n \geq n_0. \quad (1.12)$$

Condition 1.5 (Density bound for large weights).

(a) For all $R > 1$, there exist $\varepsilon_1 > 0$ and $n_1 \in \mathbb{N}$ such that for every $1 \leq x \leq R$ and $n \geq n_1$,

$$x \frac{d}{dx} \log f_n(x) \geq \varepsilon_1 s_n. \quad (1.13)$$

(b) For all $C > 1$, there exist $\varepsilon_1 > 0$ and $n_1 \in \mathbb{N}$ such that (1.13) holds for every $n \geq n_1$ and every $x \geq 1$ satisfying $f_n(x) \leq C f_n(1) \log n$.

Notice that [Condition 1.3](#) implies that $f_n(1) \sim u_n$ whenever $s_n = o(n)$. Indeed, by (1.8) we can write $u_n = f_n(x_n^{1/s_n})$ for $x_n = (-n \log(1 - 1/n))^{s_n}$. Since $s_n = o(n)$, we have $x_n = 1 - o(1)$ and the monotonicity of f_n implies that $f_n(x_n^{1/s_n})/f_n(1) \rightarrow 1$. We remark also that (1.4) remains valid if $u_n(x)$ is replaced by $f_n(x)$.

In [Section 1.3](#) the universality class is explored in more detail, and several examples that satisfy [Conditions 1.3–1.5](#) are discussed. The results from [Section 1.1](#) generalize as follows:

Theorem 1.6. Suppose that [Condition 1.5 \(a\)](#) is satisfied for a positive sequence $(s_n)_n$ with $s_n / \log \log n \rightarrow \infty$. Let $M^{(1)}, M^{(2)}$ be i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x . Then

$$f_n^{-1}(W_n) \xrightarrow{d} M^{(1)} \vee M^{(2)}. \quad (1.14)$$

[Theorem 1.6](#) is proved in [Section 7](#).

Theorem 1.7. Let $(s_n)_n$ be a positive sequence with $s_n \rightarrow \infty$. Suppose that [Condition 1.5 \(a\)](#) holds and in addition [Condition 1.4](#) holds for any value $\delta_0 \in (0, 1)$. Then, for each fixed $m \in \mathbb{N}$, the topology of the FPP smallest-weight tree to the nearest m vertices, as well as the weights along its edges, converges in distribution to invasion percolation on the Poisson-weighted infinite tree.

[Theorem 1.7](#) will be formalized in [Theorem 2.7](#).

[Theorem 1.1](#) follows from [Theorem 1.6](#). [Theorem 1.2](#) follows from [Theorem 1.7](#) because in the case $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$, (1.12) and (1.13) hold identically with $\varepsilon_1 = 1 = \varepsilon_0$. Note that neither [Theorem 1.6](#) nor [Theorem 1.7](#) requires [Condition 1.3](#).

1.3 Regularity: Sufficient condition for the universality class

Theorems 1.6 and 1.7 apply also to n -independent distributions. The following example collects some edge-weight distributions that satisfy Conditions 1.3–1.5:

Example 1.8 (Examples of weight distributions).

- (a) Let $(s_n)_{n \in \mathbb{N}} \in (0, \infty)^{\mathbb{N}}$ with $s_n \rightarrow \infty$. Take $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$, i.e., $F_Y(y) = 1 - e^{-y^{1/s_n}}$.
- (b) Let $\rho > 0, \kappa \in (0, 1)$. Take $Y_e^{(K_n)} \stackrel{d}{=} \exp(-(E/\rho)^{1/\kappa})$, for which $F_Y(y) = \exp(-\rho(\log(1/y))^\kappa)$, and define $s_n = \frac{(\log n)^{1/\kappa-1}}{\kappa \rho^{1/\kappa}}$.
- (c) Let $\rho > 0, \alpha \in (0, 1)$. Take $Y_e^{(K_n)} \stackrel{d}{=} \exp(-\rho e^{\alpha E}/\alpha)$, for which

$$F_Y(y) = \exp\left(-\frac{1}{\alpha} \log\left(\frac{\alpha}{\rho} \log(1/y)\right)\right),$$

and define $s_n = \rho n^\alpha$.

- (d) Let $\rho > 0, \alpha \in (0, 1)$. Take $Y_e^{(K_n)} \stackrel{d}{=} \exp(-\rho E^{-\alpha}/\alpha)$, for which

$$F_Y(y) = 1 - \exp\left(-\left(\frac{\alpha}{\rho} \log(1/y)\right)^{-1/\alpha}\right),$$

and define $s_n = \rho n^\alpha$.

For Example 1.8 (a) and (d), the edge weights distributions are expressed in terms of strictly increasing functions of E , namely $g(x) = x^{s_n}$ and $g(x) = \exp(-\rho x^{-\alpha}/\alpha)$ respectively. We therefore get $F_Y(y)$ by (1.7). For Example 1.8 (b) and (c), the function is strictly decreasing instead, so we get $F_Y(y)$ by the following analogue of (1.7) for the case $Y = h(E)$ and h strictly decreasing:

$$F_Y(y) = e^{-h^{-1}(y)} \quad \text{and} \quad h(x) = F_Y^{-1}(e^{-x}). \quad (1.15)$$

Indeed,

$$F_Y(y) = \mathbb{P}(Y_e^{(K_n)} \leq y) = \mathbb{P}(E_e^{(K_n)} \geq h^{-1}(y)) = e^{-h^{-1}(y)}. \quad (1.16)$$

For Example 1.8 (b) $h(x) = \exp(-(x/\rho)^{1/\kappa})$ and $h^{-1}(y) = \rho(\log(1/y))^\kappa$. By (1.15) $F_Y(y) = \exp(-\rho(\log(1/y))^\kappa)$. Analogously, for Example 1.8 (c) we have $h(x) = \exp(-\rho e^{\alpha x}/\alpha)$ and $h^{-1}(y) = \frac{1}{\alpha} \log(\frac{\alpha}{\rho} \log(1/y))$. By (1.15) we get $F_Y(y) = \exp(-\frac{1}{\alpha} \log(\frac{\alpha}{\rho} \log(1/y)))$. In Corollary 1.10 below, we derive the values s_n that were stated in Example 1.8.

Example 1.8 (b)–(d) are instances of the following family of edge-weight distributions: Suppose that the edge weights $Y_e^{(K_n)}$ follow an n -independent distribution F_Y with no atoms and $u \mapsto u(F_Y^{-1})'(u)/F_Y^{-1}(u)$ is regularly varying with index $-\alpha$ as $u \downarrow 0$, i.e.,

$$u \frac{d}{du} \log F_Y^{-1}(u) = u^{-\alpha} L(1/u) \quad \text{for all } u \in (0, 1), \quad (1.17)$$

where $t \mapsto L(t)$ is slowly varying as $t \rightarrow \infty$. The following proposition shows that edge weights of this type satisfy our conditions with $s_n = n^\alpha L(n)$:

Proposition 1.9. *For an n -independent distribution $Y_e^{(K_n)} \stackrel{d}{=} g(E)$ with distribution function F_Y , the function $u \mapsto u(F_Y^{-1})'(u)/F_Y^{-1}(u)$ is regularly varying with index $-\alpha$ as $u \downarrow 0$ if and only if $x \mapsto xg'(x)/g(x)$ is regularly varying with index $-\alpha$ as $x \downarrow 0$. If either of these equivalent conditions holds then, writing for all $x > 0$, $u \in (0, 1)$,*

$$u \frac{d}{du} \log F_Y^{-1}(u) = u^{-\alpha} L(1/u) \quad \text{and} \quad x \frac{d}{dx} \log g(x) = x^{-\alpha} \tilde{L}(1/x) \quad (1.18)$$

it holds that

$$L(t) \sim \tilde{L}(t) \quad \text{as } t \rightarrow \infty. \quad (1.19)$$

Furthermore if either of the asymptotically equivalent sequences

$$s_n = n^\alpha L(n) \quad \text{or} \quad s_n = n^\alpha \tilde{L}(n) \quad (1.20)$$

satisfies $s_n \rightarrow \infty$, then Conditions 1.3, 1.4 and 1.5 (a) hold, while if in addition $s_n / \log \log n \rightarrow \infty$ then Condition 1.5 (b) holds as well.

Note that replacing $(s_n)_n$ by an asymptotically equivalent sequence makes no difference in Conditions 1.3–1.5. Moreover, every sequence $(s_n)_n$ of the form $s_n = n^\alpha L(n)$, for $\alpha \geq 0$ and L slowly varying at infinity, can be obtained from a n -independent distribution by taking $\log F_Y^{-1}(u) = \int u^{-1-\alpha} L(1/u) du$, i.e., the indefinite integral of the function $u \mapsto u^{-1-\alpha} L(1/u)$. For a given n -independent distribution, Proposition 1.9 allows us to define the sequence s_n using either F_Y or g , whichever is more convenient. See the proof of Corollary 1.10 for details.

Proof of Proposition 1.9. The equivalence follows by noting that $\tilde{L}(1/x) = e^{-x} (\frac{x}{1-e^{-x}})^{1+\alpha} L(1/(1-e^{-x}))$ (see (1.7) and (1.18)), so that $\tilde{L}(t)$ is slowly varying as $t \rightarrow \infty$ if and only if $L(t)$ is. Conditions 1.4 and 1.5 (a) follow from the observation that (recall (1.8))

$$\frac{x f'_n(x)}{f_n(x)} = \frac{\frac{x}{n} g'(\frac{x}{n})}{g(\frac{x}{n})} = n^\alpha x^{-\alpha} \tilde{L}(n/x) \quad (1.21)$$

and the definition of slowly varying, while Condition 1.3 follows from the computation

$$\begin{aligned} \log \frac{f_n(x^{1/s_n})}{f_n(1)} &= \int_1^{x^{1/s_n}} \frac{f'_n(\xi)}{f_n(\xi)} d\xi = \int_1^x \frac{f'_n(u^{1/s_n})}{f_n(u^{1/s_n})} \frac{1}{s_n} u^{1/s_n-1} du \\ &= \int_1^x \frac{n^\alpha u^{-\alpha/s_n} \tilde{L}(n/u^{1/s_n})}{n^\alpha \tilde{L}(n)} \frac{du}{u} \rightarrow \log x, \end{aligned} \quad (1.22)$$

since $\tilde{L}(n/u^{1/s_n})/\tilde{L}(n) \rightarrow 1$ and $u^{-\alpha/s_n} \rightarrow 1$, uniformly over u in a compact subset of $(0, \infty)$.

Finally, assume $s_n / \log \log n \rightarrow \infty$ and fix any $C, R \in (1, \infty)$. By Condition 1.5 (a), there is $\varepsilon_1 = \varepsilon(R) > 0$ such that for sufficiently large n , $\log(f_n(R)/f_n(1)) \geq \varepsilon_1 s_n \int_1^R 1/\tilde{x} d\tilde{x}$. Therefore, for sufficiently large n , $f_n(R) \geq f_n(1) R^{\varepsilon_1 s_n} \geq C f_n(1) \log n$ by Condition 1.5 (a) and the inequality (1.13) holds for any value $x \geq 1$ such that $f_n(x) \leq C f_n(1) \log n$, as required by Condition 1.5 (b). \square

Corollary 1.10. *The edge-weight distributions in Example 1.8 (a)–(d), with the associated sequences $(s_n)_n$, satisfy conditions Conditions 1.3–1.5.*

Proof. For [Example 1.8 \(a\)](#) is immediate. For [Example 1.8 \(d\)](#), [Proposition 1.9](#) allows us to derive the sequence s_n . In this case it is more convenient to use the second equation in [\(1.18\)](#) with $g(x) = \exp(-\rho x^{-\alpha}/\alpha)$, thus we get

$$x \frac{d}{dx} \log g(x) = \rho x^{-\alpha} \quad \text{and} \quad \tilde{L}(1/x) = \rho. \quad (1.23)$$

We conclude that the [Conditions 1.3, 1.4](#) and [1.5 \(a\)](#) hold with $s_n = \rho n^\alpha$. Since $s_n / \log \log n \rightarrow \infty$, [Condition 1.5 \(b\)](#) holds as well.

For [Example 1.8 \(c\)](#) it is more convenient to use the first equation in [\(1.18\)](#) with $F_Y(y) = \exp(-\frac{1}{\alpha} \log(\frac{\alpha}{\rho} \log(1/y)))$. Thus $F_Y^{-1}(u) = \exp(-\frac{\alpha}{\rho} u^{-\alpha})$,

$$u \frac{d}{du} \log F_Y^{-1}(u) = \rho u^{-\alpha} \quad \text{and} \quad L(1/u) = \rho. \quad (1.24)$$

By [Proposition 1.9](#), we can set $s_n = \rho n^\alpha$, and [Conditions 1.3–1.5](#) hold since $s_n / \log \log n \rightarrow \infty$.

For [Example 1.8 \(b\)](#), we first observe that $U \stackrel{d}{=} 1 - U$ and $U \stackrel{d}{=} e^{-E}$, so $e^{-E} \stackrel{d}{=} 1 - e^{-E}$ and $E \stackrel{d}{=} -\log(1 - e^{-E})$. Thus we get

$$Y_e^{(\kappa_n)} \stackrel{d}{=} \exp(-(E/\rho)^{1/\kappa}) \stackrel{d}{=} \exp\left[-\left(-\log(1 - e^{-E})/\rho\right)^{1/\kappa}\right]. \quad (1.25)$$

In this case is more convenient to use the second equation in [\(1.18\)](#) with

$$g(x) = \exp\left[-\left(-\log(1 - e^{-x})/\rho\right)^{1/\kappa}\right], \quad (1.26)$$

thus we get

$$x \frac{d}{dx} \log g(x) = \frac{x}{\kappa \rho^{1/\kappa}} \left(-\log(1 - e^{-x})\right)^{1/\kappa-1} \frac{e^{-x}}{1 - e^{-x}} \sim \frac{1}{\kappa \rho^{1/\kappa}} (\log(1/x))^{1/\kappa-1}, \quad (1.27)$$

so that $L(n) \sim \frac{1}{\kappa \rho^{1/\kappa}} (\log n)^{1/\kappa-1}$. Asymptotic equivalence does not affect [Conditions 1.3–1.5](#), so we can set $s_n = \frac{(\log n)^{1/\kappa-1}}{\kappa \rho^{1/\kappa}}$. By [Proposition 1.9](#) and the observation that $s_n / \log \log n \rightarrow \infty$ since $\kappa < 1$, [Conditions 1.3–1.5](#) hold. \square

An example of an edge-weight distribution that is n -dependent but not $Y_e \stackrel{d}{=} E^{s_n}$ is the following:

Example 1.11. Let $(s_n)_n$ be a positive sequence with $s_n \rightarrow \infty$, $s_n = o(n^{1/3})$. Let U be a positive, continuous random variable with distribution function G and

$$\lim_{u \downarrow 0} u \frac{d}{du} \log G^{-1}(u) = 1. \quad (1.28)$$

Take $Y_e^{(\kappa_n)} \stackrel{d}{=} U^{s_n}$, i.e., $F_Y(y) = G(y^{1/s_n})$. Examples for G are the uniform distribution on an interval $(0, b)$, for any $b > 0$, or the exponential distribution with any parameter.

Lemma 1.12. *The edge weights of [Example 1.11](#) satisfy [Conditions 1.4–1.5](#), and [Condition 1.3](#) when $s_n / \log \log n \rightarrow \infty$.*

Proof. Write $U \stackrel{d}{=} h(E)$ for an increasing function $h: (0, \infty) \rightarrow (0, \infty)$, so that $f_n(x) = h(x/n)^{s_n}$ and

$$\frac{x f'_n(x)}{f_n(x)} = s_n \frac{\frac{x}{n} h'(\frac{x}{n})}{h(\frac{x}{n})}. \quad (1.29)$$

By (1.7), the assumption (1.28) on the distribution function of U is equivalent to

$$\lim_{x \downarrow 0} x \frac{d}{dx} \log h(x) = 1. \quad (1.30)$$

Conditions 1.4–1.5 follow immediately, and Condition 1.3 follows as in Proposition 1.9. \square

2 Fine results on the IP part of FPP

In this section, we argue that the optimal path between two vertices can be divided into two parts: The local neighbourhoods of the two endpoints that follow IP dynamics by Theorem 1.7, and the main part of the path which is characterized in terms of a branching process. The main results of this paper connect the maximal weight $M^{(1)}$ in IP to the transition time between these two regimes, and give a detailed description of the topology of the neighbourhood contained in the IP part.

2.1 Coupling FPP to a continuous-time branching process

To understand the random neighbourhood of a vertex in the complete graph, we study the first passage *exploration process*. Recall from (1.1) that $d_{K_n, Y^{(K_n)}}(i, j)$ denotes the total cost of the optimal path $\pi_{i,j}$ between vertex i and j . For a vertex $j \in V(K_n)$, let the *smallest-weight tree* $\text{SWT}_t^{(j)}$ be the connected subgraph of K_n defined by

$$\begin{aligned} V(\text{SWT}_t^{(j)}) &= \{i \in V(K_n) : d_{K_n, Y^{(K_n)}}(i, j) \leq t\}, \\ E(\text{SWT}_t^{(j)}) &= \{e \in E(K_n) : e \in \pi_{j,i} \text{ for some } i \in V(\text{SWT}_t^{(j)})\}. \end{aligned} \quad (2.1)$$

Note that $\text{SWT}_t^{(j)}$ is indeed a tree: if two optimal paths $\pi_{j,k}, \pi_{j,k'}$ pass through a common vertex i , both paths must contain $\pi_{j,i}$ since the minimizers of (1.1) are unique.

To visualize the process $(\text{SWT}_t^{(j)})_{t \geq 0}$, think of the edge weight $Y_e^{(K_n)}$ as the time required for fluid to flow across the edge e . Place a source of fluid at j and allow it to spread through the graph. Then $V(\text{SWT}_t^{(j)})$ is precisely the set of vertices that have been wetted by time t , while $E(\text{SWT}_t^{(j)})$ is the set of edges along which, at any time up to t , fluid has flowed from a wet vertex to a previously dry vertex. Equivalently, an edge is added to $\text{SWT}_t^{(j)}$ whenever it becomes completely wet, with the additional rule that an edge is not added if it would create a cycle.

Because fluid begins to flow across an edge only after one of its endpoints has been wetted, the *age* of a vertex – the length of time that a vertex has been wet – determines how far fluid has traveled along the adjoining edges. Given $\text{SWT}_t^{(j)}$, the future of the exploration process will therefore be influenced by the current ages of vertices in $\text{SWT}_t^{(j)}$, and the nature of this effect depends on the probability law of the edge weights $(Y_e^{(K_n)})_e$. In the sequel, for a subgraph $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ of K_n , we write \mathcal{G} instead of $V(\mathcal{G})$ for the vertex set when there is no risk of ambiguity.

To study the smallest-weight tree from a vertex, say vertex 1, let us consider the time until the first vertex is added. By construction, $\min_{i \in [n] \setminus \{1\}} Y_{\{1,i\}}^{(K_n)} \stackrel{d}{=} f_n(\frac{n}{n-1}E)$ (cf. (1.9)), where E is an

exponential random variable of mean 1. We next extend this to describe the distribution of the order statistics of the weights of edges from vertex 1 to *all* other vertices.

Denote by $Y_{(k)}^{(K_n)}$ the k^{th} smallest weight from $(Y_{\{1,i\}}^{(K_n)})_{i \in [n] \setminus \{1\}}$. Then $(Y_{(k)}^{(K_n)})_{k \in [n-1]} \stackrel{d}{=} (f_n(S_{k,n}))_{k \in [n-1]}$, where $S_{k,n} = \sum_{j=1}^k \frac{n}{n-j} E_j$ and $(E_j)_{j \in [n-1]}$ are i.i.d. exponential random variables with mean 1. The fact that the distribution of $S_{k,n}$ depends on n is awkward, and can be changed by using a *thinned* Poisson point process. Let $X_1 < X_2 < \dots$ be the points of a Poisson point process with intensity 1, so that $X_k \stackrel{d}{=} \sum_{j=1}^k E_j = \lim_{n \rightarrow \infty} S_{k,n}$. To each $k \in \mathbb{N}$, we associate a *mark* M_k which is chosen uniformly at random from $[n]$, different marks being independent. We *thin* a point X_k when $M_k = 1$ (since 1 is the initial vertex) or when $M_k = M_{k'}$ for some $k' < k$. Then

$$(Y_{(k)}^{(K_n)})_{k \in [n-1]} \stackrel{d}{=} (f_n(X_k))_{k \in \mathbb{N}, X_k \text{ unthinned}}. \quad (2.2)$$

In the next step, we extend this result to the smallest-weight tree $\text{SWT}^{(1)}$ using a relation to FPP on the Poisson-weighted infinite tree. Before giving the definitions, we recall the Ulam–Harris notation for describing trees.

Define the tree $\mathcal{T}^{(1)}$ as follows. The vertices of $\mathcal{T}^{(1)}$ are given by finite sequences of natural numbers headed by the symbol \emptyset_1 , which we write as $\emptyset_1 j_1 j_2 \dots j_k$. The sequence \emptyset_1 denotes the root vertex of $\mathcal{T}^{(1)}$. We concatenate sequences $v = \emptyset_1 i_1 \dots i_k$ and $w = \emptyset_1 j_1 \dots j_m$ to form the sequence $vw = \emptyset_1 i_1 \dots i_k j_1 \dots j_m$ of length $|vw| = |v| + |w| = k + m$. Identifying a natural number j with the corresponding sequence of length 1, the j^{th} child of a vertex v is vj , and we say that v is the parent of vj . Write $p(v)$ for the (unique) parent of $v \neq \emptyset_1$, and $p^k(v)$ for the ancestor k generations before, for $k \leq |v|$.

We can place an edge (which we could consider to be directed) between every $v \neq \emptyset_1$ and its parent; this turns $\mathcal{T}^{(1)}$ into a tree with root \emptyset_1 . With a slight abuse of notation, we will use $\mathcal{T}^{(1)}$ to mean both the set of vertices and the associated graph, with the edges given implicitly according to the above discussion, and we will extend this convention to any subset $\tau \subset \mathcal{T}^{(1)}$. We also write $\partial\tau = \{v \notin \tau : p(v) \in \tau\}$ for the set of children one generation away from τ .

The Poisson-weighted infinite tree is an infinite edge-weighted tree in which every vertex has infinitely many (ordered) children. To describe it formally, we associate weights to the edges of $\mathcal{T}^{(1)}$. By construction, we can index these edge weights by non-root vertices, writing the weights as $X = (X_v)_{v \neq \emptyset_1}$, where the weight X_v is associated to the edge between v and its parent $p(v)$. We make the convention that $X_{v_0} = 0$.

Definition 2.1 (Poisson-weighted infinite tree). The *Poisson-weighted infinite tree* (PWIT) is the random tree $(\mathcal{T}^{(1)}, X)$ for which $X_{vk} - X_{v(k-1)}$ is exponentially distributed with mean 1, independently for each $v \in \mathcal{T}^{(1)}$ and each $k \in \mathbb{N}$. Equivalently, the weights (X_{v1}, X_{v2}, \dots) are the (ordered) points of a Poisson point process of intensity 1 on $(0, \infty)$, independently for each v .

Motivated by (2.2), we study FPP on $\mathcal{T}^{(1)}$ with edge weights $(f_n(X_v))_v$:

Definition 2.2 (First passage percolation on the Poisson-weighted infinite tree). For FPP on $\mathcal{T}^{(1)}$ with edge weights $(f_n(X_v))_v$, let the FPP edge weight between $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$ and $p(v)$ be $f_n(X_v)$. The FPP distance from \emptyset_1 to $v \in \mathcal{T}^{(1)}$ is

$$T_v = \sum_{k=0}^{|v|-1} f_n(X_{p^k(v)}) \quad (2.3)$$

and the FPP exploration process $\text{BP}^{(1)} = (\text{BP}_t^{(1)})_{t \geq 0}$ on $\mathcal{T}^{(1)}$ is defined by $\text{BP}_t^{(1)} = \{v \in \mathcal{T}^{(1)} : T_v \leq t\}$.

Note that the FPP edge weights $(f_n(X_{vk}))_{k \in \mathbb{N}}$ are themselves the points of a Poisson point process on $(0, \infty)$, independently for each $v \in \mathcal{T}^{(1)}$. The intensity measure of this Poisson point process, which we denote by μ_n , is the image of Lebesgue measure on $(0, \infty)$ under f_n . Since f_n is strictly increasing by assumption, μ_n has no atoms and we may abbreviate $\mu_n((a, b])$ as $\mu_n(a, b)$ for simplicity. Thus μ_n is characterized by

$$\mu_n(a, b) = f_n^{-1}(b) - f_n^{-1}(a), \quad \int_0^\infty h(y) d\mu_n(y) = \int_0^\infty h(f_n(x)) dx, \quad (2.4)$$

for any measurable function $h: [0, \infty) \rightarrow [0, \infty)$.

Clearly, and as suggested by the notation, the FPP exploration process $\mathbf{BP}^{(1)}$ is a continuous-time branching process:

Proposition 2.3. *The process $\mathbf{BP}^{(1)}$ is a continuous time branching process (CTBP), started from a single individual \emptyset_1 , where the ages at childbearing of an individual form a Poisson point process with intensity μ_n , independently for each individual. The time T_v is the birth time $T_v = \inf \{t \geq 0: v \in \mathbf{BP}_t^{(1)}\}$ of the individual $v \in \mathcal{T}^{(1)}$.*

Similar to the analysis of the weights of the edges containing vertex 1, we now introduce a thinning procedure. Define $M_{\emptyset_1} = 1$ and to each other $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$ associate a mark M_v chosen independently and uniformly from $[n]$.

Definition 2.4 (Thinning). The root $\emptyset_1 \in \mathcal{T}^{(1)}$ is not thinned, i.e. *unthinned*. The vertex $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$ is *thinned* if it has an ancestor $v_0 = p^k(v)$ (possibly v itself) such that $M_{v_0} = M_w$ for some unthinned vertex w with $T_w < T_{v_0}$.

Note that whether or not a vertex v is thinned can be assessed recursively in terms of earlier-born vertices¹ and therefore Definition 2.4 is not circular.

Write $\widetilde{\mathbf{BP}}_t^{(1)}$ for the subgraph of $\mathbf{BP}_t^{(1)}$ consisting of unthinned vertices. If a vertex $v \in \mathcal{T}^{(1)}$ is thinned, then so are all its descendants, and this implies that $\widetilde{\mathbf{BP}}_t^{(1)}$ is a tree for all t .

Definition 2.5. Given a subset $\tau \subset \mathcal{T}^{(1)}$ and marks $M = (M_v: v \in \tau)$ with $M_v \in [n]$, define $\pi_M(\tau)$ to be the subgraph of K_n induced by the mapping $\tau \rightarrow [n]$, $v \mapsto M_v$. That is, $\pi_M(\tau)$ has vertex set $\{M_v: v \in \tau\}$, with an edge between M_v and $M_{p(v)}$ whenever $v, p(v) \in \tau$.

Note that if the marks $(M_v)_{v \in \tau}$ are distinct then $\pi_M(\tau)$ and τ are isomorphic graphs.

The following theorem establishes a close connection between FPP on K_n and FPP on the PWIT with edge weights $(f_n(X_v))_v$:

Theorem 2.6 (Coupling to FPP on PWIT). *The law of $(\mathbf{SWT}_t^{(1)})_{t \geq 0}$ is the same as the law of $(\pi_M(\widetilde{\mathbf{BP}}_t^{(1)}))_{t \geq 0}$.*

Theorem 2.6 is based on an explicit coupling between the edge weights $(Y_e^{(K_n)})_e$ on K_n and $(X_v)_v$ on $\mathcal{T}^{(1)}$. A general form of those couplings and the proof of Theorem 2.6 are given in Section 4.

¹At least up until the time $t = \sup_{x \geq 0} f_n(x)$ when $\mathbf{BP}_t^{(1)}$ ceases to be finite a.s. However, before time t , the root \emptyset_1 has had infinitely many children, a.s., so all available marks have been used and all vertices born after time t are thinned, a.s.

2.2 Relation to invasion percolation on the PWIT

Under our scaling assumptions, FPP on the PWIT is closely related to *invasion percolation* (IP) on the PWIT which is defined as follows. Set $\text{IP}^{(1)}(0)$ to be the subgraph consisting of \emptyset_1 only. For $k \in \mathbb{N}$, form $\text{IP}^{(1)}(k)$ inductively by adjoining to $\text{IP}^{(1)}(k-1)$ the boundary vertex $v \in \partial \text{IP}^{(1)}(k-1)$ of minimal weight. We note that, since we consider only the relative ordering of the various edge weights, we can use either the PWIT edge weights $(X_v)_v$ or the FPP edge weights $(f_n(X_v))_v$.

Write $\text{IP}^{(1)}(\infty) = \bigcup_{k=1}^{\infty} \text{IP}^{(1)}(k)$ for the limiting subgraph. We remark that $\text{IP}^{(1)}(\infty)$ is a strict subgraph of $\mathcal{T}^{(1)}$ a.s. (in contrast to FPP, which eventually explores every vertex). Indeed, define

$$M^{(1)} = \sup \{X_v : v \in \text{IP}^{(1)}(\infty) \setminus \{\emptyset_1\}\}, \quad (2.5)$$

the largest weight of an invaded edge. Then $\mathbb{P}(M^{(1)} < x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x , as in [Theorem 1.6](#). Indeed, the event $\{M^{(1)} < x\}$ is the event that, if we remove from $\mathcal{T}^{(1)}$ all edges of weight $X_{vk} > x$, the component of \emptyset_1 in the resulting subgraph is infinite. We remark that, a.s., the supremum in (2.5) is attained, and the unique edge of weight $M^{(1)}$ is invaded after a finite number of steps (see [Proposition 6.2](#) below).

The value $x = 1$ acts as a *critical value* for the PWIT. Indeed, if we remove all edges of weight $X_{vk} > x$, then the subtree containing the roots is a branching process with $\text{Poi}(x)$ offspring distribution. Hence for $x \leq 1$ the tree is finite a.s., while for $x > 1$ the tree is infinite with positive probability. As a result, IP on the PWIT will have to accept edges of weight $X_{vk} > 1$ infinitely often, and we have $M^{(1)} > 1$ a.s.

We next explain the connection between FPP and IP on the PWIT, under our scaling assumptions for the edge weights. We emphasize that FPP depends on n via the edge weights $(f_n(X_v))_v$, whereas IP is independent of n .

By [Condition 1.3](#), we can approximate

$$f_n(X_v) \approx f_n(1)(X_v)^{s_n}. \quad (2.6)$$

(Note that for the FPP exploration process, the only effect of multiplying by $f_n(1)$ is to rescale time.) Since the ℓ^s norm converges towards the ℓ^∞ norm (recall (1.3)), we see that, when $s_n \rightarrow \infty$, a small edge weight is almost negligible when added to a larger edge weight. Therefore, the sequence of edges added under the FPP dynamics with weights $f_n(X_v) \approx f_n(1)(X_v)^{s_n}$ can be well approximated by adding the boundary edge having the smallest weight, that is, by the IP dynamics. Moreover, the time for the FPP exploration is dominated by the time $f_n(M^{(1)})$ spent exploring the edge of *largest* weight, and until this edge has been explored only a finite number of other edges will be explored.

To formalize this discussion, consider the smallest-weight tree $\text{SWT}^{(1)}$ on K_n started from vertex 1, as defined in (2.1). Write $\tau_k = \inf\{t \geq 0 : |E(\text{SWT}_t^{(1)})| = k\}$ for the time when the k^{th} edge is added to $\text{SWT}^{(1)}$. Then we have the following local weak convergence result, formalizing [Theorem 1.2](#) and [Theorem 1.7](#):

Theorem 2.7 (Coupling to IP on the PWIT). *Suppose $\lim_{n \rightarrow \infty} f_n(x + \delta)/f_n(x) = \infty$ for each $x, \delta > 0$. Then the smallest-weight tree $\text{SWT}^{(1)}$ on K_n can be coupled to invasion percolation $\text{IP}^{(1)}$ on one copy of the PWIT such that, for any fixed $m \in \mathbb{N}$,*

$$\mathbb{P}(\text{SWT}_{\tau_k}^{(1)} = \pi_M(\text{IP}^{(1)}(k)) \text{ for all } k \leq m) = 1 - o(1). \quad (2.7)$$

Theorem 2.7 is proved in Section 4.4. The convergence in Theorem 2.7 is the local weak convergence in the sense of Benjamini and Schramm [8] for appropriately chosen metrics. Theorem 1.7 follows from Theorem 2.7 because given $x, \delta > 0$ we can apply Condition 1.4 (with $1 - \delta_0 = x$, if $x < 1$) or Condition 1.5 (a) (with $R = x + \delta$, if $x + \delta > 1$) to find a value $\varepsilon > 0$ such that for sufficiently large n ,

$$\log f_n(x + \delta) - \log f_n(x) = \int_x^{x+\delta} t \frac{d}{dt} \log f_n(t) dt/t \geq \varepsilon s_n \log(1 + \delta/x). \quad (2.8)$$

Applying the exponential function on both sides, $f_n(x + \delta)/f_n(x) \rightarrow \infty$ follows from $s_n \rightarrow \infty$.

The heuristic comparison between FPP and IP ceases to be valid when a smaller edge weight $f_n(X_v \wedge X_w)$ is no longer negligible when added to a larger edge weight $f_n(X_v \vee X_w)$. By (2.6), this requires $|X_v - X_w| = \Theta(1/s_n)$. By our discussion of the critical value, only edge weights $X_v \approx 1$ will be relevant to the large-scale behavior. It follows that once edge weights belonging to a *critical window* $[1 - \Theta(1/s_n), 1 + \Theta(1/s_n)]$ become numerous, the heuristic fails and the connection to IP on the PWIT ceases to hold. Since we are mainly interested in the case that $1/s_n \gg n^{-1/3}$ (see in particular [11]), the critical window observed here is wider than the critical window for the Erdős-Rényi random graph; cf. [6].

For IP on the PWIT, the weight of the maximal weight edge that IP uses after time $k^2 t$ scales as $1 + \mathcal{U}_t/k$, where $(\mathcal{U}_t)_{t \geq 0}$ is a limiting stochastic process (see [7, Proposition 3.3 and Theorem 1.6] and the remarks following [3, Theorem 31]). In particular, these weights become of the order $1 + \Theta(1/s_n)$ when the size of the IP cluster is $\Theta(s_n^2)$. This suggests that the maximal size of the smallest-weight tree that allows its dynamics to be coupled to IP on the PWIT is $o(s_n^2)$. However, we do not need and will not prove such a strong result.

2.3 Exploration from two sources

So far we have studied the smallest-weight tree $\text{SWT}^{(1)}$ from one vertex and its coupling to a suitably thinned version of a CTBP starting from one root. To study the optimal path between vertices 1 and 2, a standard approach would be to place sources of fluid on both vertices and wait for the two smallest-weight trees to merge. Appealing to the coupling in Theorem 2.6, equally we could study the evolution of two independent CTBPs $\text{BP}^{(1)}$ and $\text{BP}^{(2)}$ with original ancestors \emptyset_1 and \emptyset_2 . For this method it is important that the two trees grow in a similar fashion. However, the heuristics in Section 2.2 show that $\text{BP}^{(j)}$ spends a considerable amount of time, namely $f_n(M^{(j)}) \gg f_n(1)$, waiting for a single edge to be explored, during which time only finitely many other edges are explored. Since $M^{(1)} \neq M^{(2)}$ a.s., our scaling assumptions on f_n mean that the times $f_n(M^{(1)}), f_n(M^{(2)})$ will be quite different.

For this reason, we will not grow the two CTBPs at the same speed. When one of them becomes large enough (what this means precisely will be explained later), it has to wait for the other one to catch up. We call this procedure *freezing*.

To formalize this, let \mathcal{T} be the disjoint union of two independent copies $(\mathcal{T}^{(j)}, X^{(j)})$, $j \in \{1, 2\}$, of the PWIT. We shall assume that the copies $\mathcal{T}^{(j)}$ are vertex-disjoint, with roots \emptyset_j , so that we can unambiguously write X_v instead of $X_v^{(j)}$ for $v \in \mathcal{T}^{(j)}$, $v \neq \emptyset_j$. The notation introduced for $\mathcal{T}^{(1)}$ is used verbatim on \mathcal{T} . For example, for any subset $\tau \subseteq \mathcal{T}$, we write $\partial\tau = \{v \notin \tau : p(v) \in \tau\}$ for the boundary vertices of τ .

The FPP process on \mathcal{T} with edge weights $(f_n(X_v))_v$ starting from \emptyset_1 and \emptyset_2 is equivalent to the union $\text{BP} = \text{BP}^{(1)} \cup \text{BP}^{(2)}$ of two CTBPs. Let $T_{\text{fr}}^{(j)}$ be a stopping time with respect to the

filtration induced by $\text{BP}^{(j)}$, $j \in \{1, 2\}$. We call $T_{\text{fr}}^{(1)}$ and $T_{\text{fr}}^{(2)}$ *freezing times* and run BP until $T_{\text{fr}}^{(1)} \wedge T_{\text{fr}}^{(2)}$, the time when one of the two CTBPs is large enough (see [Definition 2.12](#) for the precise definition of what large enough means). Then we freeze the larger CTBP and allow the smaller one to evolve normally until it is large enough, at time $T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$. At this time, which we call the *unfreezing time* $T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$, both CTBPs resume their usual evolution. We denote by $R_j(t)$ the on-off processes describing this behavior: that is, for $j \in \{1, 2\}$,

$$R_j(t) = (t \wedge T_{\text{fr}}^{(j)}) + ((t - T_{\text{unfr}}) \vee 0). \quad (2.9)$$

The version of $\text{BP} = (\text{BP}_t)_{t \geq 0}$ including freezing is then given by

$$\mathcal{B}_t = \bigcup_{j=1}^2 \mathcal{B}_t^{(j)}, \quad \mathcal{B}_t^{(j)} = \{v \in \mathcal{T}^{(j)} : T_v \leq R_j(t)\} = \text{BP}_{R_j(t)}^{(j)} \quad \text{for all } t \geq 0. \quad (2.10)$$

As with \mathcal{T} , we can consider \mathcal{B}_t to be the union of two trees by placing an edge between each non-root vertex $v \notin \{\emptyset_1, \emptyset_2\}$ and its parent. We denote by $T_v^{\mathcal{B}} = \inf \{t \geq 0 : v \in \mathcal{B}_t\}$ the arrival time of the individual $v \in \mathcal{T}$ in $\mathcal{B} = (\mathcal{B}_t)_{t \geq 0}$. Using the left-continuous inverse of $R_j(t)$, defined by

$$R_j^{-1}(y) = \inf \{t \geq 0 : R_j(t) \geq y\} = \begin{cases} t & \text{if } t \leq T_{\text{fr}}^{(j)}, \\ T_{\text{unfr}} - T_{\text{fr}}^{(j)} + t & \text{if } t > T_{\text{fr}}^{(j)}, \end{cases} \quad (2.11)$$

we obtain $T_v^{\mathcal{B}} = R_j^{-1}(T_v)$ for $v \in \mathcal{T}^{(j)}$.

2.4 Freezing a CTBP

For very fine results about the branching process and the freezing times, we strengthen [Condition 1.5](#) to the following condition:

Condition 2.8 (Density bound for large weights). *There exist $\varepsilon_1 > 0$ and $n_1 \in \mathbb{N}$ such that*

$$x \frac{d}{dx} \log f_n(x) \geq \varepsilon_1 s_n \quad \text{for every } x \geq 1, n \geq n_1. \quad (2.12)$$

As explained, the purpose of the freezing is to guarantee a comparable growth of the two CTBPs. One requirement is therefore that the edge of weight $f_n(M^{(j)})$ is explored before freezing, for each $j = 1, 2$ (and they are instantaneously unfrozen after the last of these times). The second requirement is that the two CTBPs exhibit typical branching process dynamics. To make this precise, recall that a typical branching process grows exponentially where the growth rate is given by its Matlthusian parameter λ_n : Writing $\hat{\mu}_n(\lambda) = \int e^{-\lambda y} d\mu_n(y)$ for the Laplace transform of the intensity measure μ_n , $\lambda_n > 0$ is the unique solution to

$$\hat{\mu}_n(\lambda_n) = 1. \quad (2.13)$$

Asymptotically, λ_n scales like $1/f_n(1)$:

Lemma 2.9. *Suppose Conditions 1.3, 1.4 and 2.8 hold for a positive sequence $(s_n)_n$ with $s_n \rightarrow \infty$. Then $\lambda_n f_n(1) \rightarrow e^{-\gamma}$, where γ is Euler's constant.*

Lemma 2.9 is proved in Section 3. The same reasoning is used to prove a more general statement in [Part II, Theorem 5.3] (see [Part II, Sections 5.2 and 5.3]) but we include the proof here to avoid circularity.

Hence, we expect $\mathbf{BP}^{(j)} = (\mathbf{BP}_t^{(j)})_{t \geq 0}$ to grow exponentially at rate λ_n for large times t . However, initially, $\mathbf{BP}^{(j)}$ does not grow in this way. The reason is that the exponential growth typical of a branching process arises primarily from rare individuals that have an unusually large number of offspring in a very short time. Formally, for $v \in \mathcal{T}^{(1)}$, write $\mathbf{BP}^{(v)}$ for the branching process of descendants of v , re-rooted and time-shifted to start at $t = 0$. That is,

$$\mathbf{BP}_t^{(v)} = \{w \in \mathcal{T}^{(1)} : vw \in \mathbf{BP}_{T_v+t}^{(1)}\}. \quad (2.14)$$

For instance, $\mathbf{BP}^{(1)} = \mathbf{BP}^{(\emptyset_1)}$, and $(\mathbf{BP}^{(v)})_{p(v)=\emptyset_1}$ are independent of each other and of $(T_v)_{p(v)=\emptyset_1}$.

Definition 2.10. Given $R < \infty$, a vertex $v \in \mathcal{T}^{(j)}$ is R -*lucky* if $|\mathbf{BP}_{f_n(1)}^{(v)}| \geq Rs_n^2$.

That is, an R -lucky vertex has Rs_n^2 descendants by the time it reaches age $f_n(1)$. The following proposition states that this happens with probability at least a constant times $1/s_n$:

Proposition 2.11. Suppose Conditions 1.3, 1.4 and 2.8 hold for a positive sequence $(s_n)_n$ with $s_n \rightarrow \infty$. Fix $R \in (0, \infty)$. There exists $\delta > 0$ such that for every $r \in (0, R]$ there is some $n_0 \in \mathbb{N}$ such that $\mathbb{P}(v \text{ is } r\text{-lucky}) \geq \delta/(s_n\sqrt{r})$ for all $v \in \mathcal{T}$ and all $n \geq n_0$.

Proposition 2.11 is proved in Section 5.3.

Once an R -lucky vertex v is born (for some $R > 0$), another R -lucky vertex (for some $R > 0$) is likely to be born soon thereafter. Indeed, Condition 1.3 implies that between ages $f_n(1)$ and $2f_n(1)$, the number of new children of v will be Poisson with mean of order $1/s_n$. The same is true for the order s_n^2 initial descendants of v , so that a total of order s_n children is expected to be born during this time. Among these, of order 1 can be expected to repeat the unlikely event performed by v and thereby perpetuate the growth.

We therefore expect that $\mathbf{BP}^{(j)}$ will exhibit typical branching dynamics, with exponential growth on the time scale $f_n(1)$, only once $\mathbf{BP}^{(j)}$ is large enough so that the intensity of new births, in the time scale $f_n(1)$, is at least s_n .

This motivates the following definition of the freezing times:

Definition 2.12 (Freezing). Define, for $j = 1, 2$, the *freezing times*

$$T_{\text{fr}}^{(j)} = \inf \left\{ t \geq 0 : \sum_{v \in \mathbf{BP}_t^{(j)}} \int_{t-T_v}^{\infty} e^{-\lambda_n(y-(t-T_v))} d\mu_n(y) \geq s_n \right\}, \quad (2.15)$$

and the *unfreezing time* $T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$. The *frozen cluster* is given by

$$\mathcal{B}_{\text{fr}} = \mathcal{B}_{T_{\text{unfr}}} = \mathcal{B}_{\text{fr}}^{(1)} \cup \mathcal{B}_{\text{fr}}^{(2)} \quad \text{where} \quad \mathcal{B}_{\text{fr}}^{(j)} = \mathcal{B}_{T_{\text{fr}}^{(j)}}^{(j)}. \quad (2.16)$$

The random variable $\int_{t-T_v}^{\infty} e^{-\lambda_n(y-(t-T_v))} d\mu_n(y)$ represents the expected number of future offspring of vertex $v \in \mathbf{BP}_t^{(j)}$, exponentially time-discounted at rate λ_n . Recall from (2.9) that $R_j(t) = (t \wedge T_{\text{fr}}^{(j)}) + ((t - T_{\text{unfr}}) \vee 0)$. Thus, each CTBP evolves at rate 1 until its expected number of future offspring, exponentially time-discounted at rate λ_n , first exceeds s_n . At that time the configuration is “frozen” and ceases to evolve until both sides have been frozen. The two sides, which are now of a comparably large size, are then simultaneously unfrozen and thereafter evolve at rate 1. Henceforth we will always assume this choice of $T_{\text{fr}}^{(1)}, T_{\text{fr}}^{(2)}$.

We next investigate the asymptotics of the freezing times:

Theorem 2.13 (Scaling of the freezing times). *Suppose Conditions 1.3, 1.4 and 2.8 hold for a positive sequence $(s_n)_n$ with $s_n \rightarrow \infty$. The freezing times satisfy $f_n^{-1}(T_{\text{fr}}^{(j)}) \xrightarrow{\mathbb{P}} M^{(j)}$ for $j = 1, 2$.*

Theorem 2.13, which is proved in Section 6.2, confirms that the two CTBPs $\text{BP}^{(1)}$ and $\text{BP}^{(2)}$ are indeed large enough relatively soon after exploring the edges of weight $f_n(M^{(1)})$ and $f_n(M^{(2)})$, respectively.

The next result states that soon after unfreezing, an R -lucky vertex is born.

Lemma 2.14. *Suppose Conditions 1.3, 1.4 and 2.8 hold for a positive sequence $(s_n)_n$ with $s_n \rightarrow \infty$. Fix $R < \infty$ and let $v_{j,R}$ denote the first R -lucky vertex in $\mathcal{B}^{(j)}$ born after time T_{unfr} . Then $T_{v_{j,R}}^{\mathcal{B}} = T_{\text{unfr}} + O_{\mathbb{P}}(f_n(1))$.*

Lemma 2.14 is proved in Section 5.4.

2.5 Decomposing FPP into IP and branching dynamics

Similarly to the definition of the CTBPs with freezing \mathcal{B} , we can study the smallest-weight trees on K_n with freezing. However, in contrast to the definition of \mathcal{B} , there is competition between the two clusters and we have to actively forbid their merger. We do not need or prove any statements about the smallest-weight trees from two sources in this paper, and, therefore, give only the following informal definition. We refer to Part II for extensive results in this direction.

The two smallest-weight trees with freezing on K_n started from vertices 1 and 2 are two disjoint trees $\mathcal{S}^{(j)} = (\mathcal{S}_t^{(j)})_{t \geq 0}$, $j \in \{1, 2\}$, on K_n with root 1 and 2, respectively. Initially $\mathcal{S}^{(j)}$ contains only vertex j which can be viewed as the source of a fluid that differs from the fluid originating at j' with $\{j, j'\} = \{1, 2\}$. When a vertex becomes wet, it is added to the tree of the corresponding fluid and all adjacent edges between the new vertex and the other tree can no longer transport fluid. Moreover, fluid originating from j cannot travel between times $T_{\text{fr}}^{(j)}$ and T_{unfr} . We write $\mathcal{S}_t = \mathcal{S}_t^{(1)} \cup \mathcal{S}_t^{(2)}$ for all $t \geq 0$ call \mathcal{S}_t the smallest-weight tree from two sources.

Theorems 2.6 and 2.13 imply that the smallest-weight trees without freezing, i.e., the choice $R_1(t) = R_2(t) = t$, would behave almost like the asymmetric choices $R_1(t) = t$, $R_2(t) = 0$ or $R_1(t) = 0$, $R_2(t) = t$. Write $\mathcal{S}^{(\text{id})} = (\mathcal{S}_t^{(\text{id})})_{t \geq 0}$ for the smallest-weight tree starting from vertices 1 and 2, constructed as $\mathcal{S} = (\mathcal{S}_t)_{t \geq 0}$ but with R_j replaced by the identity map, for $j \in \{1, 2\}$ (that is, $\mathcal{S}^{(\text{id})}$ is the smallest-weight tree from two sources without freezing). Denote by $\mathcal{S}^{(\text{id},1)}$ the connected component of $\mathcal{S}^{(\text{id})}$ containing vertex 1. Our results suggest that

$$\frac{|\mathcal{S}_{\infty}^{(\text{id},1)}|}{n} \xrightarrow{d} \theta_{\text{id}} \quad \text{and} \quad \frac{|\mathcal{S}_{\infty}^{(1)}|}{n} \xrightarrow{d} \theta_{\text{fr}}, \quad (2.17)$$

where θ_{id} and θ_{fr} are random variables with $\mathbb{P}(\theta_{\text{id}} = 1) = \mathbb{P}(\theta_{\text{id}} = 0) = 1/2$ and $\mathbb{P}(\theta_{\text{fr}} \in (0, 1)) = 1$. This means that freezing is necessary to guarantee that both trees are asymptotically of comparable size.

Similarly to Theorem 2.6, we can couple \mathcal{B}_t and \mathcal{S}_t . To this end, we introduce a thinning procedure for $\mathcal{B} = (\mathcal{B}_t)_{t \geq 0}$ analogous to Definition 2.4. Define $M_{\emptyset_j} = j$, for $j = 1, 2$. To each other $v \in \mathcal{T} \setminus \{\emptyset_1, \emptyset_2\}$, we associate a mark M_v chosen uniformly and independently from $[n]$.

Definition 2.15. The vertex $v \in \mathcal{T} \setminus \{\emptyset_1, \emptyset_2\}$ is *thinned* if it has an ancestor $v_0 = p^k(v)$ (possibly v itself) such that $M_{v_0} = M_w$ for some unthinned vertex w with $T_w^{\mathcal{B}} < T_{v_0}^{\mathcal{B}}$.

As in the remarks below [Definition 2.4](#), this definition is not circular, and we write $\tilde{\mathcal{B}}_t$ for the subgraph of \mathcal{B}_t consisting of unthinned vertices.

Recall the subgraph $\pi_M(\tau)$ of K_n introduced in [Definition 2.5](#), which we extend to the case where $\tau \subset \mathcal{T}$.

Theorem 2.16. *There exists a coupling such that $\mathcal{S}_t^{(1)} \cup \mathcal{S}_t^{(2)} = \pi_M(\tilde{\mathcal{B}}_t)$ for all $t \geq 0$ almost surely.*

[Theorem 2.16](#) is formalized and proved as [Part II, Theorem 2.15].

We will not use the result of [Theorem 2.16](#) in this paper. Nevertheless, [Theorem 2.16](#) motivates our study of \mathcal{B} and freezing times because it relates FPP on the complete graph (n -independent dynamics run on an n -dependent weighted graph) with an exploration defined in terms of a pair of Poisson-weighted infinite trees (n -dependent dynamics run on an n -independent weighted graph). By analyzing the dynamics of \mathcal{B} when n and s_n are large, we obtain a fruitful *dual picture*: a *static* approximation by IP, valid when the number of explored vertices is small and independent of n ; followed by a *dynamic* rescaled branching process approximation, valid when the number of explored vertices is large, that is also essentially independent of n .

An important goal of this paper is to understand the dynamics of the IP part which is formalized as the frozen cluster. Our main results are collected in the following theorem:

Theorem 2.17 (Properties of the frozen cluster). *Suppose Conditions [1.3](#), [1.4](#) and [2.8](#) hold for a positive sequence $(s_n)_n$ with $s_n \rightarrow \infty$.*

- (a) *The volume $|\mathcal{B}_{\text{fr}}|$ of the frozen cluster is $O_{\mathbb{P}}(s_n^2)$.*
- (b) *The diameter $\max\{|v| : v \in \mathcal{B}_{\text{fr}}\}$ of the frozen cluster is $O_{\mathbb{P}}(s_n)$.*

[Theorem 2.17](#) (a) and (b) are proved in Sections [5.3](#) and [6.3](#), respectively. [Theorem 2.17](#) will allow us to ignore the elements coming from the frozen cluster in analysis of the smallest-weight path. For instance, part (b) will be used to show that path lengths within the frozen cluster are negligible compared to overall path lengths. We believe that the diameter $\max\{|v| : v \in \mathcal{B}_{\text{fr}}\}$ of the frozen cluster really is of order s_n , but we will not need this and therefore also not prove this. Since $s_n \rightarrow \infty$, this intuitively explained ‘long paths’ in our title. This analysis will be carried out in [\[11\]](#) where we are interested in global properties of the FPP process. There also the title will be substantiated, since it will be shown that H_n is of order $s_n \log(n/s_n^3)$ when $s_n = o(n^{1/3})$.

2.6 Discussion of our results

In this section we briefly discuss our results and state open problems. For a more detailed discussion of the results in this paper and in our companion paper [\[11\]](#), as well as an extensive discussion of the relations to the literature, we refer to [Part II, Section 1.4].

First passage percolation (FPP) on the complete graph is closely approximated by invasion percolation (IP) on the Poisson-weighted infinite tree (PWIT), studied in [\[3\]](#), whenever $s_n \rightarrow \infty$. See [Theorem 1.7](#) and the discussion in [Section 2.1](#). However, this relationship is a *local* one, and the scaling of s_n relative to n controls whether the two objects are *globally* comparable. [Theorem 1.6](#) shows that the weights are globally comparable provided $s_n/\log \log n \rightarrow \infty$. For the hopcount, the appropriate comparison is to the minimal spanning tree (MST) on the complete graph, obtained from running IP with a simple no-loops constraint. Path lengths in the MST scale as $n^{1/3}$ (see [\[2\]](#) and [\[1\]](#)). We conjecture that, for $s_n^3/n \rightarrow \infty$, FPP on the complete graph is in the same universality class as IP. It would be of great interest to make this connection precise by showing,

for example, that $H_n/n^{1/3}$ converges in distribution, and that the scaling limit of H_n is the same as the scaling limit of the graph distance between two uniform vertices in the MST.

The local graph convergence from [Theorem 1.7](#) and weight convergence from [Theorem 1.6](#) are the first two in a hierarchy of possible comparisons between FPP and the MST. Strengthening the previous statement about the scaling limit of H_n , we can ask whether the optimal path between vertices $i, j \in [n]$ equals (under a suitable coupling) the unique path in the MST from i to j ; whether the union of the optimal paths² from vertex i to every other vertex $j \neq i$ equals the entire MST; and whether these unions agree simultaneously for every $i \in [n]$. Assuming hypotheses similar to [Conditions 1.3–1.5](#), it would be of interest to know how s_n must grow relative to n in order for each of these events to occur.

3 Growth and density bounds for f_n and μ_n

Throughout this section, we assume [Conditions 1.4](#) and [2.8](#). Further assumptions will be stated explicitly. We will reserve the notation ε_0, δ_0 for some fixed choice of the constants in [Conditions 1.4](#) and [2.8](#), with ε_0 chosen small enough to satisfy both conditions.

The aim of the section is to explore the key implications of [Conditions 1.4](#) and [2.8](#) on f_n and on the intensity measure μ_n .

Lemma 3.1. *There exists $n_0 \in \mathbb{N}$ such that*

$$f_n(x) \leq \left(\frac{x}{x'}\right)^{\varepsilon_0 s_n} f_n(x') \quad \text{whenever } 1 - \delta_0 \leq x \leq x', n \geq n_0. \quad (3.1)$$

Proof. Divide [\(1.12\)](#) or [\(2.12\)](#) by x and integrate between x and x' to obtain $\log f_n(x') - \log f_n(x) \geq \varepsilon_0 s_n (\log x' - \log x)$ whenever $1 - \delta_0 \leq x \leq x', n \geq n_0$, as claimed. \square

We call [Condition 2.8](#) a density bound because it implies the following lemma:

Lemma 3.2. *For n sufficiently large, on the interval $(f_n(1 - \delta_0), \infty)$, the measure μ_n is absolutely continuous with respect to Lebesgue measure and*

$$\mathbb{1}_{\{y > f_n(1 - \delta_0)\}} d\mu_n(y) \leq \frac{1}{\varepsilon_0 s_n} \frac{f_n^{-1}(y)}{y} dy. \quad (3.2)$$

Proof. By [Conditions 1.4](#) and [2.8](#), f_n is strictly increasing on $(1 - \delta_0, \infty)$, so $y = f_n(\mu_n(0, y))$ for $y > f_n(1 - \delta_0)$. Differentiating and again applying [Conditions 1.4](#) and [2.8](#), we get

$$1 = f'_n(\mu_n(0, y)) \frac{d}{dy} \mu_n(0, y) \geq \varepsilon_0 s_n \frac{f_n(\mu_n(0, y))}{\mu_n(0, y)} \frac{d}{dy} \mu_n(0, y), \quad y > f_n(1 - \delta_0). \quad \square$$

Lemma 3.3. *For n sufficiently large, the density of μ_n with respect to Lebesgue measure is at most $1/(\varepsilon_0 s_n f_n(1))$ on the interval $(f_n(1), \infty)$.*

Proof. From [Lemma 3.1](#) it follows immediately that $f_n^{-1}(y) \leq (y/f_n(1))^{1/\varepsilon_0 s_n} \leq y/f_n(1)$ for all $y > f_n(1)$ and sufficiently large n . The result now follows from [Lemma 3.2](#). \square

In the lemma below, the notation $\mu(t + dy)$ denotes the translation of the measure μ by t .

²This union of optimal paths will be the smallest-weight tree $\text{SWT}_t^{(i)}$ from [Section 2](#) in the limit $t \rightarrow \infty$.

Lemma 3.4. Given $\varepsilon, \bar{\varepsilon} > 0$, there exist $n_0 \in \mathbb{N}$ and $K < \infty$ such that, for all $n \geq n_0$ and $t \geq 0$,

$$\int e^{-\varepsilon y/f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(t + dy) \leq \bar{\varepsilon}/s_n. \quad (3.3)$$

Proof. By Lemma 3.3, for large n , the density of μ_n with respect to Lebesgue measure is bounded from above by $1/(\varepsilon_0 s_n f_n(1))$ on $(f_n(1), \infty)$. Hence, for $K > 1$,

$$\int e^{-\varepsilon y/f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(t + dy) \leq \int_t^\infty e^{-\varepsilon(y-t)/f_n(1)} \mathbb{1}_{\{y-t \geq K f_n(1)\}} \frac{dy}{\varepsilon_0 s_n f_n(1)} = \frac{e^{-\varepsilon K}}{\varepsilon_0 s_n \varepsilon}. \quad \square$$

We are now in the position to prove Lemma 2.9. Recall from the discussion around (2.13) that $\hat{\mu}_n(\lambda) = \int e^{-\lambda y} d\mu_n(y)$ denotes the Laplace transform of μ_n .

Proof of Lemma 2.9. We begin by proving

$$\hat{\mu}_n \left(\frac{ae^{-\gamma}}{f_n(1)} \right) = 1 - \frac{\log a}{s_n} + o(1/s_n). \quad (3.4)$$

Recalling (2.4), we have

$$\hat{\mu}_n(\lambda) = \int_0^\infty e^{-\lambda f_n(\tilde{x})} d\tilde{x}. \quad (3.5)$$

Write $\tilde{f}_n(\tilde{x}) = f_n(\tilde{x}) \mathbb{1}_{\{\tilde{x} \geq 1-\delta_0\}}$. Then

$$\hat{\mu}_n(\lambda) = \int_0^\infty e^{-\lambda \tilde{f}_n(\tilde{x})} d\tilde{x} - \int_0^{1-\delta_0} (1 - e^{-\lambda f_n(\tilde{x})}) d\tilde{x} \quad (3.6)$$

and take $\lambda = ae^{-\gamma}/f_n(1)$ to estimate $\int_0^{1-\delta_0} (1 - e^{-ae^{-\gamma} f_n(\tilde{x})/f_n(1)}) d\tilde{x} = O(f_n(1-\delta_0)/f_n(1)) = o(1/s_n)$ by Lemma 3.1. Hence, for the purposes of proving (3.4), it is no loss of generality to assume that $f_n(x^{1/s_n}) \leq f_n(1)x^{\varepsilon_0}$ for all $x \leq 1$.

Inspired by the example $Y_e \stackrel{d}{=} E^{s_n}$ in (1.10), where $f_n(\tilde{x}) = f_n(1)\tilde{x}^{s_n}$, we compute

$$\int_0^\infty e^{-\lambda f_n(1)\tilde{x}^{s_n}} d\tilde{x} = \int_0^\infty \frac{1}{s_n} x^{1/s_n-1} e^{-\lambda f_n(1)x} dx = \left(\frac{\Gamma(1 + 1/s_n)^{s_n}}{\lambda f_n(1)} \right)^{1/s_n}. \quad (3.7)$$

In particular, setting $\lambda = a\Gamma(1 + 1/s_n)^{s_n}/f_n(1)$ gives $\int_0^\infty \exp(-a\Gamma(1 + 1/s_n)^{s_n} \tilde{x}^{s_n}) d\tilde{x} = a^{-1/s_n}$, which is $1 - (\log a)/s_n + o(1/s_n)$. Subtracting this from (3.5), we can therefore prove (3.4) if we show that

$$s_n \int_0^\infty \left(e^{-ae^{-\gamma} f_n(\tilde{x})/f_n(1)} - e^{-a\Gamma(1+1/s_n)^{s_n} \tilde{x}^{s_n}} \right) d\tilde{x} \rightarrow 0, \quad (3.8)$$

or equivalently, by the substitution $\tilde{x} = x^{1/s_n}$, if we show that

$$\int_0^\infty x^{1/s_n-1} \left(e^{-ae^{-\gamma} f_n(x^{1/s_n})/f_n(1)} - e^{-a\Gamma(1+1/s_n)^{s_n} x} \right) dx \rightarrow 0. \quad (3.9)$$

Note that $\Gamma(1 + 1/s)^s \rightarrow e^{-\gamma}$ as $s \rightarrow \infty$. Together with Condition 1.3, this implies that the integrand in (3.9) converges pointwise to 0. For $x \leq 1$, $f_n(x^{1/s_n}) \leq f_n(1)x^{\varepsilon_0}$ means that the integrand is bounded by $O(x^{\varepsilon_0-1} + 1)$. For $x \geq 1$, Lemma 3.1 implies that the integrand is bounded by $e^{-\delta x^{\varepsilon_0}}$ for some $\delta > 0$. Dominated convergence therefore completes the proof of (3.4).

To conclude Lemma 2.9 from (3.4), we use the monotonicity of μ_n . Taking $a > 1$ shows that $\hat{\mu}_n(ae^{-\gamma}/f_n(1)) < 1$ for all n large enough, implying $\lambda_n < ae^{-\gamma}/f_n(1)$ and $\limsup_{n \rightarrow \infty} \lambda_n f_n(1) \leq e^{-\gamma}$. A lower bound holds similarly. \square

Lemma 3.5. *Suppose [Condition 1.3](#) holds. Given $K < \infty$, there exist $\varepsilon_K > 0$ and $n_0 \in \mathbb{N}$ such that, for $0 \leq t \leq Kf_n(1)$ and $n \geq n_0$,*

$$\int_0^\infty e^{-\lambda_n y} \mu_n(t + dy) \geq \varepsilon_K / s_n. \quad (3.10)$$

Proof. For any $0 \leq t \leq Kf_n(1)$,

$$\int_0^\infty e^{-\lambda_n y} \mu_n(t + dy) = \int_0^\infty e^{-\lambda_n(y-t)} \mathbb{1}_{\{y \geq t\}} d\mu_n(y) \geq e^{-2\lambda_n Kf_n(1)} \mu_n(Kf_n(1), 2Kf_n(1)). \quad (3.11)$$

By [Lemma 2.9](#), $\lambda_n f_n(1)$ converges to a finite constant. By [Condition 1.3](#), $f_n(1 + x/s_n)/f_n(1) \rightarrow e^x$, and it follows that $\mu_n(Kf_n(1), 2Kf_n(1)) = f_n^{-1}(2Kf_n(1)) - f_n^{-1}(Kf_n(1)) \sim (\log 2)/s_n$. \square

4 Coupling K_n and the PWIT

In [Theorem 2.16](#), we indicated that two random processes, the first passage exploration processes \mathcal{S} and \mathcal{B} on K_n and \mathcal{T} , respectively, could be coupled. In [Section 2.1](#) we have intuitively described the coupling between FPP on K_n and on the PWIT. In this section we explain how this coupling arises as a special case of a general family of couplings between K_n , understood as a random edge-weighted graph with i.i.d. exponential edge weights, and the PWIT. In [Section 4.2](#) we define the minimal rule processes and the thinning in this context. In [Section 4.3](#) and [Section 4.4](#) we prove [Theorems 2.6](#) and [2.7](#), respectively.

4.1 Exploration processes and the definition of the coupling

As in [Sections 2.1](#) and [2.5](#), we define $M_{\partial_j} = j$, for $j = 1, 2$, and to each other $v \in \mathcal{T} \setminus \{\partial_1, \partial_2\}$, we associate a mark M_v chosen uniformly and independently from $[n]$. We next define what an exploration process is:

Definition 4.1 (Exploration process on two PWITs). Let \mathcal{F}_0 be a σ -field containing all null sets, and let (\mathcal{T}, X) be independent of \mathcal{F}_0 . We call a sequence $\mathcal{E} = (\mathcal{E}_k)_{k \in \mathbb{N}_0}$ of subsets of \mathcal{T} an *exploration process* if, with probability 1, $\mathcal{E}_0 = \{\partial_1, \partial_2\}$ and, for every $k \in \mathbb{N}$, either $\mathcal{E}_k = \mathcal{E}_{k-1}$ or else \mathcal{E}_k is formed by adjoining to \mathcal{E}_{k-1} a previously unexplored child $v_k \in \partial \mathcal{E}_{k-1}$, where the choice of v_k depends only on the weights X_w and marks M_w for vertices $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$ and on events in \mathcal{F}_0 .

Examples for exploration processes are given by FPP and IP on \mathcal{T} . For FPP, as defined in [Definition 2.2](#), it is necessary to convert to discrete time by observing the branching process at those moments when a new vertex is added, similar to [Theorem 2.7](#). The standard IP on \mathcal{T} is defined as follows. Set $\text{IP}(0) = \{\partial_1, \partial_2\}$. For $k \in \mathbb{N}$, form $\text{IP}(k)$ inductively by adjoining to $\text{IP}(k-1)$ the boundary vertex $v \in \partial \text{IP}(k-1)$ of minimal weight. However, an exploration process is also obtained when we specify at each step (in any suitably measurable way) whether to perform an invasion step in $\mathcal{T}^{(1)}$ or in $\mathcal{T}^{(2)}$.

For $k \in \mathbb{N}$, let \mathcal{F}_k be the σ -field generated by \mathcal{F}_0 together with the weights X_w and marks M_w for vertices $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$. Note that the requirement on the choice of v_k in [Definition 4.1](#) can be expressed as the requirement that \mathcal{E} is $(\mathcal{F}_k)_k$ -adapted.

For $v \in \mathcal{T}$, define the exploration time of v by

$$N_v = \inf \{k \in \mathbb{N}_0 : v \in \mathcal{E}_k\}. \quad (4.1)$$

Definition 4.2 (Thinning). The vertex $v \in \mathcal{T} \setminus \{\emptyset_1, \emptyset_2\}$ is *thinned* if it has an ancestor $v_0 = p^k(v)$ (possibly v itself) such that $M_{v_0} = M_w$ for some unthinned vertex w with $N_w < N_{v_0}$. Write $\tilde{\mathcal{E}}_k$ for the subgraph of \mathcal{E}_k consisting of unthinned vertices.

Recall the remark below [Definition 2.4](#) that explains that the definition above is not circular. We define the stopping times

$$N(i) = \inf \left\{ k \in \mathbb{N}_0 : M_v = i \text{ for some } v \in \tilde{\mathcal{E}}_k \right\}. \quad (4.2)$$

at which $i \in [n]$ first appears as a mark in the unthinned exploration process. Note that, on the event $\{N(i) < \infty\}$, $\tilde{\mathcal{E}}_k$ contains a *unique* vertex in \mathcal{T} whose mark is i , for any $k \geq N(i)$; call that vertex $V(i)$. On this event, we define

$$X(i, i') = \min_{w \in \mathcal{T}} \{X_w : M_w = i', p(w) = V(i)\}. \quad (4.3)$$

Lemma 4.3. *Conditional on $\mathcal{F}_{N(i)}$, and on the event $\{N(i) < \infty\}$, the distribution of $X(i, i')$ is exponential with mean n , independently for every i' . Moreover, $X(i, i')$ is $\mathcal{F}_{N(i)+1}$ measurable.*

Proof. The event $\{N(i) = k, V(i) = v\}$ is measurable with respect to the σ -field generated by \mathcal{F}_0 together with all edge weights X_v, X_w and marks M_v, M_w for which w is not a descendant of v . On the other hand, on $\{V(i) = v\}$, $X(i, i') = \min \{X_w : M_w = i', p(w) = v\}$ depends only on the marks and edge weights of children of v . Therefore, the distribution of $X(i, i')$ and the independence for different i' follow from the thinning property of Poisson point processes. Since the marks and edge weights of the children of $V(i)$ are measurable with respect to $\mathcal{F}_{N(i)+1}$, $X(i, i')$ is measurable with respect to this σ -field. \square

We define, for an edge $\{i, i'\} \in E(K_n)$,

$$X_{\{i, i'\}}^{(K_n)} = \begin{cases} \frac{1}{n} X(i, i') & \text{if } N(i) < N(i'), \\ \frac{1}{n} X(i', i) & \text{if } N(i') < N(i), \\ E_{\{i, i'\}} & \text{if } N(i) = N(i') = \infty \text{ or } N(i) = N(i') = 0, \end{cases} \quad (4.4)$$

where $(E_e)_{e \in E(K_n)}$ are exponential variables with mean 1, independent of each other and of $(X_v)_v$.

Theorem 4.4. *If \mathcal{E} is an exploration process on the union \mathcal{T} of two PWITs, then the edge weights $X_e^{(K_n)}$ defined in (4.4) are exponential with mean 1, independently for each $e \in E(K_n)$.*

The idea underlying [Theorem 4.4](#) is that, by [Lemma 4.3](#), each variable $\frac{1}{n} X(i, i')$ is exponentially distributed conditional on the past up to the moment $N(i)$ when it may be used to set the value of $X_{\{i, i'\}}^{(K_n)}$. However, formalizing this notion requires careful attention to the relative order of the stopping times $N(i)$ and to which $N(i)$ are infinite.

Proof. Let $(h_e)_{e \in E(K_n)}$ be an arbitrary collection of bounded, measurable functions, and abbreviate $\langle h_e \rangle = \mathbb{E}[h_e(E)]$, where E is exponential with mean 1. It suffices to prove that

$$\mathbb{E} \left(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \right) = \prod_{e \in E(K_n)} \langle h_e \rangle. \quad (4.5)$$

We proceed by induction. To begin, we partition (4.5) according to the number $\ell \in \{0, 1, \dots, n-2\}$ of indices $i \neq 1, 2$ for which $N(i) = \infty$, as well as the relative order of the finite values of

$N(i)$. Define $i_1 = 1$, $i_2 = 2$ and, given $\ell \in [n-2]$ and $\mathbf{i} = (i_3, \dots, i_{n-\ell})$, abbreviate $S_{\ell, \mathbf{i}} = [n] \setminus \{i_1, \dots, i_{n-\ell}\}$.

Note that, on the event $\{N(i) = \infty \forall i \in S_{\ell, \mathbf{i}}\}$, we have $X_{\{i, j\}}^{(K_n)} = E_{\{i, j\}}$ for $\{i, j\} \subset S_{\ell, \mathbf{i}}$ by (4.4). The $E_{\{i, j\}}$ are exponential, independently from everything else, so we may perform the integration over these variables separately. We conclude that

$$\begin{aligned} \mathbb{E}\left(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)})\right) &= \sum_{(i_3, \dots, i_n)} \mathbb{E}\left(\mathbb{1}_{\{N(i_3) < \dots < N(i_{n-1}) < N(i_n) \leq \infty\}} \prod_{e \in E(K_n)} h_e(X_e^{(K_n)})\right) \\ &+ \sum_{\ell=2}^{n-2} \sum_{\mathbf{i}=(i_3, \dots, i_{n-\ell})} \mathbb{E}\left(\mathbb{1}_{\{N(i_3) < \dots < N(i_{n-\ell}) < \infty, N(i) = \infty \forall i \in S_{\ell, \mathbf{i}}\}} \prod_{\{i, j\} \not\subset S_{\ell, \mathbf{i}}} h_{\{i, j\}}(X_{\{i, j\}}^{(K_n)})\right) \prod_{\{i, j\} \subset S_{\ell, \mathbf{i}}} \langle h_{\{i, j\}} \rangle, \end{aligned} \quad (4.6)$$

where the first sum corresponds to $\ell = 0$ and $\ell = 1$. The sums are over vectors of distinct indices $i_3, \dots, i_n \in \{3, \dots, n\}$ and $i_3, \dots, i_{n-\ell} \in \{3, \dots, n\}$, respectively, and the notation $\{i, j\} \not\subset S_{\ell, \mathbf{i}}$ means that $\{i, j\}$ is an edge with at least one endpoint in $[n] \setminus S_{\ell, \mathbf{i}} = \{i_1, \dots, i_{n-\ell}\}$.

In general, for $\mathbf{i} = (i_3, \dots, i_{n-\ell})$ given, define the events

$$A_{\ell, \mathbf{i}} = \{N(i_3) < \dots < N(i_{n-\ell}) < \infty\}, \quad B_{\ell, \mathbf{i}} = \{N(i) > N(i_{n-\ell}) \forall i \in S_{\ell, \mathbf{i}}\}. \quad (4.7)$$

We claim that, for all $\ell_0 \in [n-2]$,

$$\begin{aligned} \mathbb{E}\left(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)})\right) &= \sum_{\mathbf{i}=(i_3, \dots, i_{n-\ell_0})} \mathbb{E}\left(\mathbb{1}_{A_{\ell_0, \mathbf{i}}} \mathbb{1}_{B_{\ell_0, \mathbf{i}}} \prod_{\{i, j\} \not\subset S_{\ell_0, \mathbf{i}}} h_{\{i, j\}}(X_{\{i, j\}}^{(K_n)})\right) \prod_{\{i, j\} \subset S_{\ell_0, \mathbf{i}}} \langle h_{\{i, j\}} \rangle \\ &+ \sum_{\ell=\ell_0+1}^{n-2} \sum_{\mathbf{i}=(i_3, \dots, i_{n-\ell})} \mathbb{E}\left(\mathbb{1}_{A_{\ell, \mathbf{i}}} \mathbb{1}_{\{N(i) = \infty \forall i \in S_{\ell, \mathbf{i}}\}} \prod_{\{i, j\} \not\subset S_{\ell, \mathbf{i}}} h_{\{i, j\}}(X_{\{i, j\}}^{(K_n)})\right) \prod_{\{i, j\} \subset S_{\ell, \mathbf{i}}} \langle h_{\{i, j\}} \rangle. \end{aligned} \quad (4.8)$$

When $\ell = n-2$, by convention, the second sum vanishes, while in the first sum \mathbf{i} is the empty sequence.

The case $\ell_0 = 1$ reduces to (4.6): then $S_{\ell_0, \mathbf{i}}$ contains only one element, which we called i_n but which is in fact uniquely determined by the values i_3, \dots, i_{n-1} , and the product $\prod_{\{i, j\} \subset S_{\ell_0, \mathbf{i}}} \langle h_{\{i, j\}} \rangle$ is empty. This initializes the induction hypothesis.

We remark that in the right-hand sides of (4.6) and (4.8), the indicators already allow us to determine which of the three cases from (4.4) occurs. For notational simplicity, we will introduce this information gradually as we proceed.

Now suppose (4.8) has been proved for a given $\ell_0 < n-2$. In the first summand of the right-hand side of (4.8), we condition on $\mathcal{F}_{N(i_{n-\ell_0})}$. By Lemma 4.3 and the presence of the indicators, each factor $h_{\{i, j\}}(X_{\{i, j\}}^{(K_n)})$ is equal to a factor $h_{\{i, j\}}(\frac{1}{n}X(i, j))$ (or $h_{\{i, j\}}(\frac{1}{n}X(j, i))$, if $N(j) < N(i)$) that is $\mathcal{F}_{N(i_{n-\ell_0})}$ -measurable, with the exception of the factors $h_{\{i_{n-\ell_0}, j\}}(\frac{1}{n}X(i_{n-\ell_0}, j))$ for $j \in S_{\ell_0, \mathbf{i}}$, which are conditionally independent given $\mathcal{F}_{N(i_{n-\ell_0})}$, again by Lemma 4.3. Note furthermore that $A_{\ell_0, \mathbf{i}}, B_{\ell_0, \mathbf{i}} \in \mathcal{F}_{N(i_{n-\ell_0})}$, $A_{\ell_0, \mathbf{i}} = A_{\ell_0+1, \mathbf{i}} \cap \{N(i_{n-\ell_0-1}) < N(i_{n-\ell_0}) < \infty\}$ and $S_{\ell_0, \mathbf{i}} \cup \{i_{n-\ell_0}\} = S_{\ell_0+1, \mathbf{i}}$.

Thus,

$$\begin{aligned}
& \mathbb{E} \left(\mathbb{1}_{A_{\ell_0, \mathbf{i}}} \mathbb{1}_{B_{\ell_0, \mathbf{i}}} \prod_{\{i,j\} \not\subset S_{\ell_0, \mathbf{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \right) \prod_{\{i,j\} \subset S_{\ell_0, \mathbf{i}}} \langle h_{\{i,j\}} \rangle \\
&= \mathbb{E} \left(\mathbb{1}_{A_{\ell_0, \mathbf{i}}} \mathbb{1}_{\{N(i) > N(i_{n-\ell_0}) \forall i \in S_{\ell_0, \mathbf{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0, \mathbf{i}} \cup \{i_{n-\ell_0}\}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \right) \prod_{\{i,j\} \subset S_{\ell_0, \mathbf{i}} \cup \{i_{n-\ell_0}\}} \langle h_{\{i,j\}} \rangle \\
&= \mathbb{E} \left(\mathbb{1}_{A_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{\{N(i_{n-\ell_0-1}) < N(i_{n-\ell_0}) < N(i) \forall i \in S_{\ell_0, \mathbf{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0+1, \mathbf{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \right) \prod_{\{i,j\} \subset S_{\ell_0+1, \mathbf{i}}} \langle h_{\{i,j\}} \rangle. \quad (4.9)
\end{aligned}$$

Leaving $i_3, \dots, i_{n-\ell_0-1}$ fixed, we now sum (4.9) over all $i_{n-\ell_0} \in [n] \setminus \{i_1, \dots, i_{n-\ell_0-1}\}$. This rewrites the first sum in (4.8) as

$$\sum_{\mathbf{i}=(i_3, \dots, i_{n-\ell_0-1})} \mathbb{E} \left(\mathbb{1}_{A_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{B_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{\{N(i) < \infty \text{ for some } i \in S_{\ell_0+1, \mathbf{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0+1, \mathbf{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \right) \prod_{\{i,j\} \subset S_{\ell_0+1, \mathbf{i}}} \langle h_{\{i,j\}} \rangle. \quad (4.10)$$

However, (4.10) combines with the summand $\ell = \ell_0 + 1$ from the second sum in (4.8) (since we can rewrite $\mathbb{1}_{A_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{\{N(i) = \infty \forall i \in S_{\ell_0+1, \mathbf{i}}\}}$ as $\mathbb{1}_{A_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{B_{\ell_0+1, \mathbf{i}}} \mathbb{1}_{\{N(i) = \infty \forall i \in S_{\ell_0+1, \mathbf{i}}\}}$) to produce the first sum in (4.8) for ℓ_0 replaced by $\ell_0 + 1$. This advances the induction hypothesis, and thus completes the proof of (4.8) for all $\ell_0 \leq n - 2$.

We therefore conclude that (4.8) holds when $\ell_0 = n - 2$. In this case the second sum vanishes, while in the first sum \mathbf{i} is the empty sequence, $S_{n-2, \mathbf{i}} = [n] \setminus \{1, 2\}$ and the events $A_{n-2, \mathbf{i}}, B_{n-2, \mathbf{i}}$ always occur (note that $N(i_2) = 0$). Therefore

$$\begin{aligned}
\mathbb{E} \left(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \right) &= \mathbb{E} \left(\prod_{\{i,j\} \cap \{1,2\} \neq \emptyset} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \right) \prod_{\{i,j\} \cap \{1,2\} = \emptyset} \langle h_{\{i,j\}} \rangle \\
&= \mathbb{E} \left(h_{\{1,2\}}(E_{\{1,2\}}) \prod_{i=3}^n h_{\{1,i\}}(\tfrac{1}{n}X(1,i)) h_{\{2,i\}}(\tfrac{1}{n}X(2,i)) \right) \prod_{\{i,j\} \cap \{1,2\} = \emptyset} \langle h_{\{i,j\}} \rangle. \quad (4.11)
\end{aligned}$$

By Lemma 4.3, $(\frac{1}{n}X(1,i))_{i \geq 3}$ and $(\frac{1}{n}X(2,i))_{i \geq 3}$ are each families of independent exponential random variables with mean 1. Moreover they are mutually independent, since they are determined from the independent Poisson point processes of edge weights corresponding to \emptyset_1 and \emptyset_2 , respectively. Since furthermore $E_{\{1,2\}}$ is independent of everything, we conclude that (4.5) holds. \square

4.2 Minimal-rule exploration processes

An important class of exploration processes, which includes both FPP and IP, are those exploration processes determined by a minimal rule in the following sense:

Definition 4.5. A *minimal rule* for an exploration process \mathcal{E} on \mathcal{T} is an $(\mathcal{F}_k)_k$ -adapted sequence $(S_k, \prec_k)_{k=1}^\infty$, where $S_k \subset \partial \mathcal{E}_{k-1}$ is a (possibly empty) subset of the boundary vertices of \mathcal{E}_{k-1} and \prec_k is a strict total ordering of the elements of S_k (if any) such that the implication

$$w \in S_k, p(v) = p(w), M_v = M_w, X_v < X_w \implies v \in S_k, v \prec_k w \quad (4.12)$$

holds. An exploration process is *determined by the minimal rule* $(S_k, \prec_k)_{k=1}^\infty$ if $\mathcal{E}_k = \mathcal{E}_{k-1}$ whenever $S_k = \emptyset$ and otherwise \mathcal{E}_k is formed by adjoining to \mathcal{E}_{k-1} the unique vertex $v_k \in S_k$ that is minimal with respect to \prec_k .

In words, in every step k there is a set of boundary vertices S_k from which we can select for the next exploration step. The content of (4.12) is that, whenever a vertex $w \in S_k$ is available for selection, then all siblings of w with the same mark but smaller weight are also available for selection and are preferred over w .

For FPP without freezing on \mathcal{T} with edge weights $f_n(X_v)$, we take $v \prec_k w$ if and only if $T_v < T_w$ (recall (2.3)) and take $S_k = \partial \mathcal{E}_{k-1}$. For IP on \mathcal{T} , we have $v \prec_k w$ if and only if $X_v < X_w$; the choice of subset S_k can be used to enforce, for instance, whether the k^{th} step is taken in $\mathcal{T}^{(1)}$ or $\mathcal{T}^{(2)}$.

Recall the subtree $\tilde{\mathcal{E}}_k$ of unthinned vertices from Definition 4.2 and the subgraph $\pi_M(\tilde{\mathcal{E}}_k)$ from Definition 2.5. That is, $\pi_M(\tilde{\mathcal{E}}_k)$ is the union of two trees with roots 1 and 2, respectively, and for $v \in \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\}$, $\pi_M(\tilde{\mathcal{E}}_k)$ contains vertices M_v and $M_{p(v)}$ and the edge $\{M_v, M_{p(v)}\}$.

For any $i \in [n]$ for which $N(i) < \infty$, recall that $V(i)$ is the unique vertex of $\tilde{\mathcal{E}}_k$ ($k \geq N(i)$) for which $M_{V(i)} = i$. Define $V(i, i')$ to be the first child of $V(i)$ with mark i' .

Recalling (4.3), an equivalent characterization of $V(i, i')$ is

$$X(i, i') = X_{V(i, i')}. \quad (4.13)$$

The following lemma shows that, for an exploration process determined by a minimal rule, unthinned vertices must have the form $V(i, i')$:

Lemma 4.6. *Suppose \mathcal{E} is an exploration process determined by a minimal rule $(S_k, \prec_k)_{k=1}^\infty$ and $k \in \mathbb{N}$ is such that $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$. Let $i_k = M_{p(v_k)}$ and $i'_k = M_{v_k}$. Then $v_k = V(i_k, i'_k)$.*

Proof. By construction, $p(v_k) \in \tilde{\mathcal{E}}_{k-1}$ and $M_{p(v_k)} = i_k$, so $V(i_k) = p(v_k)$ by definition. Moreover, $V(i_k) = p(V(i_k, i'_k))$ and $M_{V(i_k, i'_k)} = i'_k = M_{v_k}$. Suppose to the contrary that $V(i_k, i'_k) \neq v_k$. By the definition of $V(i_k, i'_k)$, it follows that $X_{V(i_k, i'_k)} < X_{v_k}$ and (4.12) yields $V(i_k, i'_k) \in S_k$ and $V(i_k, i'_k) \prec_k v_k$, a contradiction since v_k must be minimal for \prec_k . \square

If \mathcal{E} is an exploration process determined by a minimal rule, then we define

$$S_k^{(K_n)} = \left\{ \{i, i'\} \in E(K_n) : i \in \pi_M(\tilde{\mathcal{E}}_{k-1}), i' \notin \pi_M(\tilde{\mathcal{E}}_{k-1}), V(i, i') \in S_k \right\} \quad (4.14)$$

and

$$e_1 \tilde{\prec}_k e_2 \iff V(i_1, i'_1) \prec_k V(i_2, i'_2), \quad e_1, e_2 \in S_k^{(K_n)}, \quad (4.15)$$

where $e_j = \{i_j, i'_j\}$ and $i_j \in \pi_M(\tilde{\mathcal{E}}_{k-1}), i'_j \notin \pi_M(\tilde{\mathcal{E}}_{k-1})$ as in (4.14).

Proposition 4.7 (Thinned minimal rule). *Suppose \mathcal{E} is an exploration process determined by a minimal rule $(S_k, \prec_k)_{k=1}^\infty$. Then, under the edge-weight coupling (4.4), the edge weights of $\pi_M(\tilde{\mathcal{E}}_k)$ are determined by*

$$X_{\{M_v, M_{p(v)}\}}^{(K_n)} = \frac{1}{n} X_v \quad \text{for any } v \in \bigcup_{k=1}^\infty \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\} \quad (4.16)$$

and generally

$$X_{\{i, i'\}}^{(K_n)} = \frac{1}{n} X_{V(i, i')} \quad \text{whenever } i \in \pi_M(\tilde{\mathcal{E}}_{k-1}), i' \notin \pi_M(\tilde{\mathcal{E}}_{k-1}) \text{ for some } k \in \mathbb{N}. \quad (4.17)$$

Moreover, for any $k \in \mathbb{N}$ for which $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$, $\pi_M(\tilde{\mathcal{E}}_k)$ is formed by adjoining to $\pi_M(\tilde{\mathcal{E}}_{k-1})$ the unique edge $e_k \in S_k^{(K_n)}$ that is minimal with respect to $\tilde{\prec}_k$.

Proposition 4.7 asserts that the subgraph $\pi_M(\tilde{\mathcal{E}}_k)$ of K_n , equipped with the edge weights $(X_e^{(K_n)})_{e \in E(\pi_M(\tilde{\mathcal{E}}_k))}$, is isomorphic as an edge-weighted graph to the subgraph $\tilde{\mathcal{E}}_k$ of \mathcal{T} , equipped with the rescaled edge weights $(\frac{1}{n}X_v)_{v \in \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\}}$. Furthermore, the subgraphs $\pi_M(\tilde{\mathcal{E}}_k)$ can be grown by an inductive rule. Thus the induced subgraphs $(\pi_M(\tilde{\mathcal{E}}_k))_{k=0}^\infty$ themselves form a minimal-rule exploration process on K_n , with a minimal rule derived from that of \mathcal{E} , with the caveat that $\tilde{\prec}_k$ may depend on edge weights from $\mathcal{E}_{k-1} \setminus \tilde{\mathcal{E}}_{k-1}$ as well as from $\pi_M(\tilde{\mathcal{E}}_{k-1})$.

Proof of Proposition 4.7. We first prove (4.17). By assumption, $N(i) \leq k-1 < N(i')$, so (4.17) is simply the first case in (4.4) (see also (4.13)).

Take $v \in \cup_{k=1}^\infty \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\}$, and assume that $v = v_k$, i.e., set $k = N_v \geq 1$. Set $i_k = M_{p(v_k)}$ and $i'_k = M_{v_k}$. By construction, $i_k \in \pi_M(\tilde{\mathcal{E}}_{k-1})$ but $i'_k \notin \pi_M(\tilde{\mathcal{E}}_{k-1})$, and according to Lemma 4.6, $v_k = V(i_k, i'_k)$. So (4.16) is a special case of (4.17).

By construction, $\pi_M(\tilde{\mathcal{E}}_k)$ is formed by adjoining to $\pi_M(\tilde{\mathcal{E}}_{k-1})$ the vertex $i_k = M_{v_k} \in [n]$ via the edge $e_k = \{i_k, i'_k\}$. By Lemma 4.6, $v_k = V(i_k, i'_k)$, and by the definition of minimal rule, the vertex v_k belongs to S_k and is minimal for \prec_k . It follows from the definitions (4.14)–(4.15) that $e_k \in S_k^{(K_n)}$ is minimal for $\tilde{\prec}_k$. \square

4.3 Coupling $\text{SWT}^{(1)}$ and $\text{BP}^{(1)}$: Proof of Theorem 2.6

In this section, we prove Theorem 2.6: that is, we couple the smallest-weight tree $\text{SWT}^{(1)}$ on K_n to a single branching process $\text{BP}^{(1)}$ on $\mathcal{T}^{(1)}$. Since this statement is concerned with processes starting from only one source, we use exploration processes on $\mathcal{T}^{(1)}$ instead of \mathcal{T} . All results from Sections 4.1 and 4.2 carry over up to obvious changes (indeed, the results hold for any finite number of copies of the PWIT).

Proof of Theorem 2.6. Let the exploration process $\mathcal{E}^{(\text{FPP}, 1)}$ on $\mathcal{T}^{(1)}$ be determined by the minimal rule where $S_k = \partial \mathcal{E}_{k-1}^{(\text{FPP}, 1)}$ and $v \prec_k^{\text{FPP}} w$ if and only if $T_v < T_w$. In words, $\mathcal{E}^{(\text{FPP}, 1)}$ performs FPP steps on $\mathcal{T}^{(1)}$ in discrete time, and it is easy to verify that $\mathcal{E}_k^{(\text{FPP}, 1)} = \text{BP}_{T_{v_k}}^{(1)}$, where v_k was defined in Definition 4.5. Throughout this proof, we choose $\mathcal{E} = \mathcal{E}^{(\text{FPP}, 1)}$.

The smallest-weight tree $\text{SWT}^{(1)}$ evolves in discrete time as follows. At time 0, $\text{SWT}_0^{(1)}$ contains only vertex 1 and no edges. Let $\tau_{k'-1}$ be the time that the $(k'-1)^{\text{st}}$ vertex, not including vertex 1, is added. Write $\tau_0 = 0$. Starting from time $\tau_{k'-1}$, the next edge added is the minimizer $e_{k'}'$ of $d_{K_n, Y^{(K_n)}}(1, i) + Y_e^{(K_n)}$ over the set of boundary edges $e = \{i, j\}$ with $i \in \text{SWT}_{\tau_{k'-1}}^{(1)}$, $j \notin \text{SWT}_{\tau_{k'-1}}^{(1)}$, and $e_{k'} = \{i_k', j_k'\}$ is added at time $\tau_{k'} = d_{K_n, Y^{(K_n)}}(1, i_{k'}') + Y_{e_{k'}}^{(K_n)}$. It is easy to verify by induction that for any $i \in \text{SWT}_{\tau_{k'-1}}^{(1)}$, $d_{K_n, Y^{(K_n)}}(1, i)$ equals the sum of edge weights $Y_e^{(K_n)}$ for e belonging to the unique path in $\text{SWT}_{\tau_{k'-1}}^{(1)}$ from 1 to i .

Define the edge weights on K_n according to (4.4) using the exploration process $\mathcal{E} = \mathcal{E}^{(\text{FPP}, 1)}$. By Theorem 4.4, the edge weights $X_e^{(K_n)}$ are i.i.d. exponential with mean 1. Hence, as discussed in Section 1.2, the edge weights $Y_e^{(K_n)} = g(X_e^{(K_n)})$ have the distribution function F_Y , so that $\text{SWT}^{(1)}$ has the correct law under this coupling.

Both $\widetilde{\text{BP}}^{(1)}$ and $\text{SWT}^{(1)}$ are increasing jump processes and $\pi_M(\widetilde{\text{BP}}_0^{(1)}) = \text{SWT}_0^{(1)}$. By an inductive argument, it suffices to show that if k, k' are such that $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$ and $\pi_M(\tilde{\mathcal{E}}_{k-1}) = \text{SWT}_{\tau_{k'-1}}^{(1)}$ then (a) the edge $e_{k'}'$ next added to $\text{SWT}_{\tau_{k'-1}}^{(1)}$ is the same as the edge $e_k = \{i_k, i_k'\}$ that is minimal with respect to $\tilde{\prec}_k$ and therefore next added to $\pi_M(\tilde{\mathcal{E}}_{k-1})$; and (b) $\tau_{k'} = T_{v_k}$. (We remark that $k > k'$ is possible.)

Let $i \in V(\pi_M(\tilde{\mathcal{E}}_{k-1}))$. The unique path in $\text{SWT}_{\tau_{k'-1}}^{(1)} = \pi_M(\tilde{\mathcal{E}}_{k-1})$ from i to 1 is the image of the unique path in $\tilde{\mathcal{E}}_{k-1}$ from $V(i)$ to \emptyset_1 under the mapping $v \mapsto M_v$ (recall [Definition 2.5](#)). According to (4.16), (1.6) and (1.8), the edge weights along this path are $Y_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)} = g(X_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)}) = g(\frac{1}{n}X_{p^{m-1}(V(i))})$ and $f_n(X_{p^{m-1}(V(i))})$, for $m = 1, \dots, |V(i)|$. Summing gives $d_{K_n, Y^{(K_n)}}(1, i) = T_{V(i)}$.

In addition, let $i' \notin V(\pi_M(\tilde{\mathcal{E}}_{k-1}))$ and write $e = \{i, i'\}$. By (4.17), $X_e^{(K_n)} = \frac{1}{n}X_{V(i, i')}$, so that $Y_e^{(K_n)} = f_n(X_{V(i, i')})$. Thus $e_{k'}$ is the minimizer of $d_{K_n, Y^{(K_n)}}(1, i) + Y_e^{(K_n)} = T_{V(i)} + f_n(X_{V(i, i')}) = T_{V(i, i')}$. By [Proposition 4.7](#), so is e_k ; thus (a) follows since the minimizer is unique. Moreover $\tau_{k'}$ is the corresponding minimum value, namely $\tau_{k'} = T_{V(i_k, i'_k)}$. According to [Lemma 4.6](#), $v_k = V(i_k, i'_k)$ and (b) follows. \square

4.4 Comparing FPP and IP: Proof of [Theorem 2.7](#)

In this section, we prove [Theorem 2.7](#) by comparing the FPP and IP dynamics on the PWIT.

Proof of [Theorem 2.7](#). It is easy to see that $\text{IP}^{(1)}$ is an exploration process determined by a minimal rule. For instance, we may take $S_k = \partial \text{IP}_{k-1}^{(1)}$ and $v \prec_k w$ if and only if $X_v < X_w$.

In fact, it will be more convenient to use a different characterization of $\text{IP}^{(1)}$. For $v \in \mathcal{T}^{(1)}$, write $\mathcal{O}(v) = (X_{(v,1)}, \dots, X_{(v,|v|)})$ for the vector of edge weights $X_{v'}$ along the path from \emptyset_1 to v , ordered from largest to smallest. Set $v \prec_k^{\text{IP}} w$ if and only if $\mathcal{O}(v)$ is lexicographically smaller than $\mathcal{O}(w)$. It is an elementary exercise that this minimal rule $(S_k, \prec_k^{\text{IP}})_{k=1}^\infty$ also determines $\text{IP}^{(1)}$.

Couple the edge weights on K_n according to (4.4), where the exploration process is $\mathcal{E}^{(\text{FPP}, 1)}$ from the proof of [Theorem 2.6](#). Fix $m \in \mathbb{N}$. With high probability, none of the first m vertices explored by $\mathcal{E}^{(\text{FPP}, 1)}$ is thinned. By [Theorem 2.6](#), it therefore suffices to show that $(\mathcal{E}_k^{(\text{FPP}, 1)})_{k=1}^m = (\text{IP}^{(1)}(k))_{k=1}^m$ whp.

Let $\varepsilon > 0$ be given. Write B_m for the collection of all vertices of the form $\emptyset_1 j_1 \dots j_r$ with $1 \leq r \leq m$ and $j_1, \dots, j_r \leq m$. (That is, B_m consists of all vertices in $\mathcal{T}^{(1)}$ within m generations for which each ancestor is at most the m^{th} child of its parent.) Note that the first m explored vertices v_1, \dots, v_m necessarily belong to B_m , for both $\mathcal{E}^{(\text{FPP}, 1)}$ and $\text{IP}^{(1)}$. Let $\delta > 0$ and write A_δ for the event that $\inf \{X_v : v \in B_m\} \geq \delta$, $\sup \{X_v : v \in B_m\} \leq 1/\delta$, and $\inf \{|X_v - X_w| : v, w \in B_m, v \neq w\} \geq \delta$. We may choose $\delta > 0$ sufficiently small that $\mathbb{P}(A_\delta) \geq 1 - \varepsilon$.

Choose $x_0 < x_1 < \dots < x_N$ such that $x_0 = \delta$, $x_N = 1/\delta$ and $x_j - x_{j-1} \leq \delta/2$ for all $j \in [N]$. By assumption, there is an $n_0 \in \mathbb{N}$ such that $f_n(x_j)/f_n(x_{j-1}) > m$ for all $j \in [N]$ and $n \geq n_0$. Hence, for any $x, x' \in [\delta, 1/\delta]$ with $x' \geq x + \delta$, the monotonicity of f_n implies

$$\frac{f_n(x')}{f_n(x)} \geq \frac{f_n(x_j)}{f_n(x_{j-1})} > m, \quad (4.18)$$

since we may choose j such that $[x_{j-1}, x_j] \subset [x, x']$. From now on assume $n \geq n_0$.

Consider any $v, w \in B_m$ such that $v \neq w$ and neither v nor w is an ancestor of the other. Then there is a smallest index j with $X_{(v,j)} \neq X_{(w,j)}$. If $X_{(v,j)} < X_{(w,j)}$ then, on A_δ ,

$$\sum_{i=j}^{|v|} f_n(X_{(v,i)}) \leq m f_n(X_{(v,j)}) < f_n(X_{(w,j)}) \leq \sum_{i=j}^{|w|} f_n(X_{(w,i)}), \quad (4.19)$$

and similarly if $X_{(v,j)} > X_{(w,j)}$. Hence $v \prec_k^{\text{FPP}} w$ if and only if $\mathcal{O}(v)$ is lexicographically smaller than $\mathcal{O}(w)$, i.e. $v \prec_k^{\text{IP}} w$, for any of the vertices v, w that may be relevant to $(\mathcal{E}_k^{(\text{FPP}, 1)})_{k=1}^m$ or

$(\text{IP}^{(1)}(k))_{k=1}^m$. Since $\mathcal{E}_0^{(\text{FPP},1)} = \text{IP}^{(1)}(0)$, it follows that, on A_δ for n sufficiently large, we have $(\mathcal{E}_k^{(\text{FPP},1)})_{k=1}^m = (\text{IP}^{(1)}(k))_{k=1}^m$, and since $\mathbb{P}(A_\delta) \geq 1 - \varepsilon$ with $\varepsilon > 0$ arbitrary, this completes the proof. \square

5 Poisson Galton-Watson trees, lucky vertices and the emergence of the frozen cluster

In this section, we prove [Proposition 2.11](#), [Lemma 2.14](#) and [Theorem 2.13](#). We begin with preliminary results on Poisson Galton-Watson trees.

Throughout this section, we assume [Conditions 1.3](#), [1.4](#) and [2.8](#). We will reserve the notation ε_0, δ_0 for some fixed choice of the constants in [Conditions 1.4](#) and [2.8](#), with ε_0 chosen small enough to satisfy both conditions.

5.1 Properties of Poisson Galton-Watson trees

Proposition 5.1. *Let $(\tau; \mathbb{P}_m)$ be a Poisson Galton-Watson tree with offspring mean $m > 0$ and write $\theta(m) = \mathbb{P}_m(|\tau| = \infty)$ for its survival probability. For $d \in (0, \infty)$, we denote by $(\tau^{(\leq d)}; \mathbb{P}_m)$ and $(\tau^{(\geq d)}; \mathbb{P}_m)$ the subgraph of τ consisting of the vertices within distance d and with distance at least d , respectively, from the root. In addition, let $(s_n)_n$ be a positive sequence with $s_n \rightarrow \infty$.*

(a) $\theta: (0, \infty) \rightarrow [0, 1]$ is non-decreasing with $\theta(m) = 0$ for $m \leq 1$, $\theta(m) > 0$ for $m > 1$, and

$$1 - \theta(m) = e^{-m\theta(m)} \quad \text{for all } m \in (0, \infty). \quad (5.1)$$

(b) As $m \downarrow 1$,

$$\theta(m) \sim 2(m-1) \quad \text{and} \quad 1 - m(1 - \theta(m)) \sim m-1, \quad (5.2)$$

and, uniformly over $m \geq 1$, $\theta(m) = O(m-1)$ and $1 - m(1 - \theta(m)) = O(m-1)$.

(c) For $m > 1$, $\mathbb{P}_m(|\tau| \in \cdot \mid |\tau| < \infty) = \mathbb{P}_{\hat{m}}(\tau \in \cdot)$, where $\hat{m} = m(1 - \theta(m)) < 1$ and

$$1 - \hat{m} \sim m-1 \quad \text{as } m \downarrow 1. \quad (5.3)$$

(d) Uniformly over $m \in (0, \infty)$,

$$\mathbb{P}_m(|\tau| = n) = \frac{1}{m\sqrt{2\pi n^3}} e^{-(m-1-\log m)n} (1 + o(1)). \quad (5.4)$$

(e) Let $K, R \in (0, \infty)$. There exists $c \in (0, \infty)$ such that for every $r \in (0, R]$ there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$\mathbb{P}_{1-K/s_n}(rs_n^2 \leq |\tau| \leq 2rs_n^2) \geq \frac{c}{s_n\sqrt{r}}. \quad (5.5)$$

(f) Let $K, R \in (0, \infty)$. There exist $n_0 \in \mathbb{N}$ and $\delta_1 > 0$ such that for all $r \in (0, R]$ and $n \geq n_0$,

$$\mathbb{P}_{1-K/s_n}(|\tau^{(\leq s_n)}| \geq rs_n^2) \geq \delta_1 \mathbb{P}_{1-K/s_n}(rs_n^2 \leq |\tau| \leq 2rs_n^2). \quad (5.6)$$

(g) Uniformly in $m \in (1, \infty)$,

$$\mathbb{P}_m(|\tau^{(\geq s_n)}| \geq 1) \leq \theta(m) + O(1/s_n). \quad (5.7)$$

Proof. (a) Identity (5.1) is obtained by considering the individuals in the first generation.

(b) From (5.1), we obtain (5.2). Since $\theta(m) \leq 1$, the uniform bounds follow.

(c) See for example Theorem 3.15 in [16]. Asymptotic (5.3) follows from (5.2).

(d) It is well-known (see for example the first display on page 951 in [3]) that

$$\mathbb{P}_m(|\tau| = n) = \frac{e^{-mn}(mn)^{n-1}}{n!}. \quad (5.8)$$

Stirling's formula yields the claim.

(e) Choose n_0 so large that $1 + o(1)$ in (5.4) is bounded from below by $1/2$ and that $-K/s_n - \log(1 - K/s_n) \leq K^2/s_n^2$ and $\lceil rs_n^2 \rceil \leq \lfloor 2rs_n^2 \rfloor$ for $n \geq n_0$. According to (d),

$$\begin{aligned} \mathbb{P}_{1-K/s_n}(rs_n^2 \leq |\tau| \leq 2rs_n^2) &\geq \sum_{i=\lceil rs_n^2 \rceil}^{\lfloor 2rs_n^2 \rfloor} \frac{1}{(1 - K/s_n)\sqrt{2\pi i^3}} e^{-K^2 i/s_n^2} \frac{1}{2} \\ &\geq \frac{1}{2\sqrt{2^4\pi r^3 s_n^6}} \frac{s_n^2}{K^2} \left(e^{-K^2 \lceil rs_n^2 \rceil / s_n^2} - e^{-K^2 \lfloor 2rs_n^2 \rfloor / s_n^2} \right). \end{aligned} \quad (5.9)$$

(f) Denote by $(\mathcal{G}_i, \mathbb{P})$ the uniform labelled rooted tree on i nodes after the labels of the children have been discarded and by $h(\mathcal{G}_i)$ the height of \mathcal{G}_i . Given $d \in (0, \infty)$ and $\varepsilon > 0$, we write $\mathcal{G}_i^{(\leq d)}$ for the subtree of \mathcal{G}_i consisting of vertices within distance d from the root and $\mathcal{G}_i \varepsilon$ for \mathcal{G}_i with distances rescaled by ε . For all $m > 0$ and $i \in \mathbb{N}$, the Poisson Galton–Watson tree with offspring mean m conditioned on having i nodes has the same distribution as \mathcal{G}_i . Moreover, $h(\mathcal{G}_i)/\sqrt{i}$ converges, as $i \rightarrow \infty$, to the maximum of $2B$ where $B = (B_t)_{t \in [0,1]}$ is a standard Brownian excursion [4, 5]. We deduce that

$$\begin{aligned} \mathbb{P}_{1-K/s_n}(|\tau^{(\leq s_n)}| \geq rs_n^2) &= \sum_{i=\lceil rs_n^2 \rceil}^{\infty} \mathbb{P}(|\mathcal{G}_i^{(\leq s_n)}| \geq rs_n^2) \mathbb{P}_{1-K/s_n}(|\tau| = i) \\ &\geq \min_{rs_n^2 \leq i \leq 2rs_n^2} \mathbb{P}(h(\mathcal{G}_i) \leq s_n) \mathbb{P}_{1-K/s_n}(rs_n^2 \leq |\tau| \leq 2rs_n^2). \end{aligned} \quad (5.10)$$

The first factor on the right-hand side of (5.10) can be estimated, for sufficiently large n , by

$$\begin{aligned} \min_{rs_n^2 \leq i \leq 2rs_n^2} \mathbb{P}\left(h(\mathcal{G}_i)/\sqrt{i} \leq s_n/\sqrt{i}\right) &\geq \min_{rs_n^2 \leq i \leq 2rs_n^2} \mathbb{P}\left(h(\mathcal{G}_i)/\sqrt{i} \leq 1/\sqrt{2r}\right) \\ &\geq \frac{1}{2} \mathbb{P}\left(\max_{t \in [0,1]} 2B_t \leq 1/\sqrt{2R}\right) = \delta_1. \end{aligned} \quad (5.11)$$

(g) Using (c), we obtain

$$\begin{aligned} \mathbb{P}_m(|\tau^{(\geq s_n)}| \geq 1) &= \mathbb{P}_m(|\tau| = \infty) + \mathbb{P}_m(|\tau| < \infty) \mathbb{P}_m(|\tau^{(\geq s_n)}| \geq 1 \mid |\tau| < \infty) \\ &= \theta(m) + (1 - \theta(m)) \mathbb{P}_{\hat{m}}(|\tau^{(\geq s_n)}| \geq 1) \\ &\leq \theta(m) + \mathbb{P}_1(|\tau^{(\geq s_n)}| \geq 1). \end{aligned} \quad (5.12)$$

The claim now follows from a standard result on critical Galton–Watson processes (see for example [15, Lemma I.10.1]). \square

5.2 The probability of luckiness: proof of Proposition 2.11

In the proof of Proposition 2.11 we use the following estimates:

Lemma 5.2. *Suppose $0 \leq X \leq Y$ are non-negative random variables such that $\mathbb{E}(X|Y) \geq pY$ a.s. Then $\mathbb{P}(X \geq m) \geq \frac{1}{2}p\mathbb{P}(Y \geq 2m/p)$ for any $m \in [0, \infty)$.*

Proof. Markov's inequality applied to $Y - X$ gives $\mathbb{P}(Y - X > (1 - \frac{1}{2}p)Y | Y) \leq (1 - p)/(1 - \frac{1}{2}p)$, so $\mathbb{P}(X \geq \frac{1}{2}pY | Y) \geq \frac{1}{2}p/(1 - \frac{1}{2}p) \geq \frac{1}{2}p$ and the result follows. \square

Lemma 5.3. *There exists $n_0 \in \mathbb{N}$ such that for any $m \in (0, \infty)$, $a \in (0, 1]$ and $n \geq n_0$,*

$$\int_0^1 f_n(amx)dx \leq \frac{f_n(1)(1 - \delta_0)^{\varepsilon_0 s_n}}{am} + \frac{f_n(m)a^{\varepsilon_0 s_n}}{\varepsilon_0 s_n} \quad (5.13)$$

Proof. By Lemma 3.1, for sufficiently large n ,

$$\begin{aligned} \int_0^1 f_n(amx)dx &\leq \frac{f_n(1 - \delta_0)}{am} + \int_{1 - \delta_0}^{am} f_n(y) \frac{dy}{am} \\ &\leq \frac{f_n(1)(1 - \delta_0)^{\varepsilon_0 s_n}}{am} + \int_0^{am} \left(\frac{y}{m}\right)^{\varepsilon_0 s_n} f_n(m) \frac{dy}{am}. \end{aligned} \quad \square$$

Proof of Proposition 2.11. Applying Lemma 5.3 with $m = 1$ and $a = 1 - K/s_n$, we find that K can be chosen such that for sufficiently large n ,

$$\int_0^1 f_n((1 - K/s_n)x) dx \leq \frac{f_n(1)}{2s_n}. \quad (5.14)$$

Let τ be the subtree of \mathcal{T} consisting of v and those descendants joined to v by a path whose edge weights X_w all satisfy $X_w \leq 1 - K/s_n$. We consider τ as a rooted labelled tree equipped with edge weights, but with the vertex labels from the PWIT forgotten³. Then, ignoring weights, τ is equal in distribution to a Poisson Galton–Watson tree with offspring mean $1 - K/s_n$ and by Proposition 5.1 (e) there is some $c > 0$ independent of $r \in (0, R]$ such that for every $r \in (0, R]$ there is $n_0 \in \mathbb{N}$ with $\mathbb{P}(4rs_n^2 \leq |\tau| \leq 8rs_n^2) \geq \frac{c}{s_n\sqrt{r}}$ for all $n \geq n_0$. Write τ' for the subtree of τ consisting of vertices within distance s_n from the root. By Proposition 5.1 (f), there is some $\delta_1 > 0$ independent of $r \in (0, R]$ such that with $\delta = \delta_1 c/4$, $\mathbb{P}(|\tau'| \geq 4rs_n^2) \geq \frac{4\delta}{s_n\sqrt{r}}$ for n sufficiently large. Conditional on τ' , the PWIT edge weights X_{wk} , $wk \in \tau'$, are uniformly distributed on $[0, 1 - K/s_n]$, and therefore the first passage edge weights $Y_{wk} = f_n(X_{wk})$ satisfy

$$\mathbb{E}(Y_{wk} | wk \in \tau') = \int_0^1 f_n((1 - K/s_n)x) dx. \quad (5.15)$$

By the definition of τ' , the graph distance between $u \in \tau'$ and v is at most s_n . Hence, (5.15) and (5.14) give $\mathbb{E}(T_u - T_v | u \in \tau') \leq s_n \frac{f_n(1)}{2s_n}$. By Markov's inequality $\mathbb{P}(T_u - T_v > f_n(1) | u \in \tau') \leq \frac{1}{2}$.

Let τ'' denote the subtree $\{u \in \tau' : T_u - T_v \leq f_n(1)\}$, so that $\mathbb{P}(u \in \tau'' | u \in \tau') \geq \frac{1}{2}$ and $\mathbb{E}(|\tau''| | |\tau'|) \geq \frac{1}{2} |\tau'|$. By Lemma 5.2 with $X = |\tau''|$, $Y = |\tau'|$, we obtain

$$\mathbb{P}\left(|\mathbf{BP}_{f_n(1)}^{(v)}| \geq rs_n^2\right) \geq \mathbb{P}(|\tau''| \geq rs_n^2) \geq \frac{1}{4} \mathbb{P}(|\tau'| \geq 4rs_n^2) \geq \frac{\delta}{s_n\sqrt{r}}. \quad \square$$

³Formally, we should consider instead of τ the subset $\tilde{\tau}$ where we replace each vertex $w \in \tau \setminus \{v\}$ by an arbitrary label $\ell(w)$ drawn independently from some continuous distribution. By a slight abuse of notation, we will refer to τ and X_w , $w \in \tau \setminus \{v\}$ instead of $\tilde{\tau}$ and $X_{\ell^{-1}(w)}$, $w \in \tilde{\tau}$. This procedure avoids the complication, implicit in our Ulam–Harris notation, that the vertex $w = \varnothing_j k_1 k_2 \dots k_r \in \mathcal{T}^{(j)}$ automatically gives information about the number of its siblings with smaller edge weights.

In the proof of [Theorem 2.17 \(a\)](#), we will bound the number of vertices of large age in the frozen cluster. To do so, we will argue that for each vertex of large age, there is an independent chance to have an R -lucky child v , where R is chosen large enough that v and its descendants are on their own sufficient to bring about the freezing time. This motivates the following definition:

Definition 5.4. Let ε_1 denote the constant from [Lemma 3.5](#) for $K = 1$. Call a vertex $v \in \mathcal{T}^{(j)}$ *lucky* if it is $(1/\varepsilon_1)$ -lucky, and set

$$T_{\text{lucky}}^{(j)} = \inf \left\{ T_v : v \in \mathcal{T}^{(j)} \setminus \{\emptyset_j\} \text{ is lucky and } T_v > T_{p(v)} + f_n(1) \right\}, \quad (5.16)$$

the first time that a lucky vertex is born to a parent of age greater than $f_n(1)$.

In view of [Definition 2.12](#) and [Lemma 3.5](#), we have

$$v \in \mathcal{T}^{(j)} \text{ is lucky} \implies T_{\text{fr}}^{(j)} \leq T_v + f_n(1). \quad (5.17)$$

In other words, a lucky vertex has enough descendants in time $f_n(1)$ that the integral in the definition (2.15) of the freezing time must be at least s_n .

Lemma 5.5. *The distribution of*

$$\sum_{v \in \text{BP}^{(j)}_{T_{\text{lucky}}^{(j)}}} (f_n^{-1}(T_{\text{lucky}}^{(j)} - T_v) - 1)^+ \quad (5.18)$$

is exponential with rate $\mathbb{P}(v \text{ is lucky})$.

Since $T_v - T_{p(v)} = f_n(X_v)$, the condition $T_v > T_{p(v)} + f_n(1)$ in the definition of $T_{\text{lucky}}^{(j)}$ is equivalent to $X_v > 1$. On the other hand, the event $\{v \text{ is lucky}\}$ depends only on the evolution of $\text{BP}^{(v)}$ until time $f_n(1)$ and is therefore determined by those descendants v' of v for which $X_{v'} \leq 1$. Because these two conditions on edge weights are mutually exclusive, it will follow that $T_{\text{lucky}}^{(j)}$ is the first arrival time of a certain Cox process. We now formalize this intuition, which requires some care.

Proof of Lemma 5.5. To avoid complications arising from our Ulam–Harris notation, we consider instead of a vertex $v = \emptyset_j k_1 k_2 \dots k_r$ the modified vertex $w(v) = \emptyset_j X_{k_1} X_{k_1 k_2} \dots X_{k_1 k_2 \dots k_r}$ formed out of the edge weights along the path from \emptyset_j to v . We can extend our usual notation for parents, length, concatenation, edge weight, and birth times to vertices of the form $w = \emptyset_j x_1 \dots x_r$, $x_i \in (0, \infty)$: for instance, $|w| = r$, $X_w = x_r$ and $T_w = f_n(x_1) + \dots + f_n(x_r)$.

Form the point measure $\mathcal{M} = \sum_{v \in \mathcal{T}} \delta_{w(v)}$ on $\cup_{r=0}^{\infty} \{\emptyset_1, \emptyset_2\} \times (0, \infty)^r$. Given \mathcal{M} , we can recover the PWIT (\mathcal{T}, X) : for instance, $X_{\emptyset_j k_1} = \inf \{x > 0 : \mathcal{M}(\{\emptyset_j\} \times (0, x]) \geq k_1\}$ and $X_{\emptyset_j k_1 k_2} = \inf \{x > 0 : \mathcal{M}(\{\emptyset_j\} \times \{X_{\emptyset_j k_1}\} \times (0, x]) \geq k_2\}$. The point measure \mathcal{M} has the advantage that a value such as $\mathcal{M}(\{\emptyset_j\} \times (a, b))$ (the number of children of \emptyset_j with edge weights in the interval (a, b)) does not reveal information about the number of sibling edges of smaller edge weight.

The Poisson property of the PWIT can be expressed in terms of \mathcal{M} by saying that, conditional on the restriction $\mathcal{M}|_{\{\emptyset_1, \emptyset_2\} \times (0, \infty)^r}$ to the first r generations, the $(r+1)^{\text{st}}$ generation $\mathcal{M}|_{\{\emptyset_1, \emptyset_2\} \times (0, \infty)^{r+1}}$ is formed as a Cox process with intensity $\mathcal{M}|_{\{\emptyset_1, \emptyset_2\} \times (0, \infty)^r} \otimes \mathbb{1}_{\{x > 0\}} dx$, where $\mathbb{1}_{\{x > 0\}} dx$ denotes Lebesgue measure on $(0, \infty)$.

We next rearrange the information contained in \mathcal{M} . Given $w = \emptyset_j x_1 \dots x_r$, let $\mathcal{M}^{(w)} = \sum_{w' : \mathcal{M}(\{ww'\})=1} \delta_{w'}$ denote the point measure corresponding to all descendants of w (thus $\mathcal{M}^{(w)} = 0$

if $\mathcal{M}(\{w\}) = 0$, and the point measures $\mathcal{M}^{(w(v))}$, $v \in \mathcal{T}$, are identically distributed and non-zero)⁴. Further, write $\mathcal{M}_{\leq}^{(w)}$ for the restriction of $\mathcal{M}^{(w)}$ to $\cup_{k=0}^{\infty} (0, 1]^k$. Since luckiness only depends on descendants explored within time $f_n(1)$, it follows that whether or not v is lucky can be determined solely in terms of $\mathcal{M}_{\leq}^{(w(v))}$. Indeed, call a point measure \mathbf{m} on $\cup_{k=0}^{\infty} (0, 1]^k$ *lucky* if $\mathbf{m}(\{w': T_{w'} \leq f_n(1)\}) \geq s_n^2/\varepsilon_1$; then v is lucky if and only if $\mathcal{M}_{\leq}^{(w(v))}$ is lucky.

Define $A^{(j)}$ to be the collection of vertices $w(v)$, $v \in \mathcal{T}^{(j)}$, that are born before time $T_{\text{lucky}}^{(j)}$. That is, for any non-root ancestor v' of v , including v itself, it is *not* the case that $T_{v'} > T_{p(v')} + f_n(1)$ and v' is lucky. To study $A^{(j)}$, we decompose the PWIT according to vertices v' that are born late (i.e., $T_{v'} > T_{p(v')} + f_n(1)$) and keep track of their early explored descendants (i.e., $\mathcal{M}_{\leq}^{(w(v'))}$).

For $w = \emptyset_j x_1 \dots x_r$, let $i_1 < \dots < i_k$ denote those indices (if any) for which $x_i > 1$, and write $w_{\ell} = \emptyset_j x_1 \dots x_{i_{\ell}}$, $\ell = 1, \dots, k$. Set $q(w, \mathcal{M})$ to be the sequence $\mathcal{M}_{\leq}^{(\emptyset_j)} \mathcal{M}_{\leq}^{(w_1)} \dots \mathcal{M}_{\leq}^{(w_k)}$. Define the rearranged point measure

$$\mathcal{R} = \int_{\{\emptyset_1, \emptyset_2\} \cup \cup_{k=1}^{\infty} \{\emptyset_1, \emptyset_2\} \times (0, \infty)^{k-1} \times (1, \infty)} \delta_{(w, q(w, \mathcal{M}))} d\mathcal{M}(w). \quad (5.19)$$

(That is, \mathcal{R} is a point measure on pairs (w, q) such that w satisfies $w \in \{\emptyset_1, \emptyset_2\}$ or $X_w > 1$, and q is a sequence of measures on $\cup_{r=0}^{\infty} (0, 1]^r$. Considering \mathcal{R} instead of \mathcal{M} corresponds to partitioning vertices according to their most recent ancestor (if any) having edge weight greater than 1.)

The Poisson property of the PWIT implies that, conditional on the restriction $\mathcal{R}|_{\{(w, q): |w| \leq r\}}$, the restriction $\mathcal{R}|_{\{(w, q): |w| = r+1\}}$ forms a Cox process with intensity measure

$$\int_{\{(w, q): |w| \leq r\}} d\mathcal{R}(w, q) \int_{\{w': |ww'| = r\}} d\mathcal{M}_{\leq}^{(w)}(w') [\delta_{ww'} \otimes \mathbb{1}_{\{x > 1\}} dx] \otimes [\delta_q \otimes d\mathbb{P}(\mathcal{M}_{\leq}^{(\emptyset_j)} \in \cdot)], \quad (5.20)$$

where $\delta_{ww'} \otimes \mathbb{1}_{\{x > 1\}} dx$ means the image of Lebesgue measure on $(1, \infty)$ under the concatenation mapping $x \mapsto ww'x$. The formula (5.20) expresses the fact that every vertex in the $(r+1)^{\text{st}}$ generation has a parent uniquely written as ww' , with $(w, q(w, \mathcal{M}))$ corresponding to a point mass in \mathcal{R} , w' corresponding to a point mass in the last entry $\mathcal{M}_{\leq}^{(w)}$ of the sequence $q(w, \mathcal{M})$, and $r = |w| + |w'|$.

Now, it is easy to verify that $A^{(j)}$ is measurable with respect to the restriction of \mathcal{R} to pairs (w, q) such that $q = \mathbf{m}_0 \dots \mathbf{m}_k$ with \mathbf{m}_{ℓ} not lucky for each $\ell \neq 0$.

Because of the Poisson property of the PWIT, as expressed via \mathcal{R} in (5.20), it follows that, conditional on $A^{(j)}$, the point measure $L = \sum_{w'' \notin A^{(j)}, p(w'') \in A^{(j)}} \delta_{T_{w''}}$ forms a Cox process on $(0, \infty)$. Furthermore, the first point of L is precisely $T_{\text{lucky}}^{(j)}$. To determine the intensity measure of L , we note that for a vertex $ww' \in A^{(j)}$, $w'' = ww'x$ satisfies $w'' \notin A^{(j)}$ if and only if $X_{w''} > 1$ and $\mathcal{M}_{\leq}^{(w'')}$ is lucky. Furthermore, the condition $T_{w''} = T_{ww'} + f_n(X_{w''}) \leq t$ is equivalent to $X_{w''} \leq f_n^{-1}(t - T_{ww'})$. Using (5.20), it follows that the cumulative intensity measure of L on $(0, t]$ is given by

$$\int_{\left\{ \begin{array}{l} (w, q): q = \mathbf{m}_0 \dots \mathbf{m}_k, \\ \mathbf{m}_{\ell} \text{ not lucky for any } \ell \neq 0 \end{array} \right\}} d\mathcal{R}(w, q) \int_{\{w': T_{ww'} \leq t\}} d\mathcal{M}_{\leq}^{(w)}(w') (f_n^{-1}(t - T_{ww'}) - 1)^+ \cdot \mathbb{P}(\mathcal{M}_{\leq}^{(\emptyset_j)} \text{ is lucky}). \quad (5.21)$$

The vertices ww' from the integral in (5.21) are in one-to-one correspondence with the vertices $u \in A^{(j)}$ satisfying $T_u \leq t$. Consequently we may re-write the cumulative intensity as

$$\sum_{u \in A^{(j)}: T_u \leq t} (f_n^{-1}(t - T_u) - 1)^+ \cdot \mathbb{P}(u \text{ is lucky}). \quad (5.22)$$

⁴In this instance, we assume that w' does not have a \emptyset_j symbol for convenience of notation.

Finally we note that $\mathbb{P}(v \text{ is lucky})$ times the sum in (5.18) is exactly the sum in (5.22) evaluated at $t = T_{\text{lucky}}^{(j)}$. (The vertex v for which $T_v = T_{\text{lucky}}^{(j)}$ does not contribute to (5.18).) The cumulative intensity in (5.22) is a.s. continuous as a function of t (since f_n^{-1} is continuous and the jumps at the times T_u are zero). But for any Cox process with continuous cumulative intensity function and infinite total intensity, it is elementary to verify that when the cumulative intensity is evaluated at the first point of the Cox process, the result is exponential with mean 1. This completes the proof. \square

5.3 Volume of the frozen cluster: proof of Theorem 2.17 (a)

To study the frozen cluster, we introduce the frozen intensity measures

$$d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \mathbb{1}_{\{y \geq 0\}} \mu_n(T_{\text{fr}}^{(j)} - T_v + dy). \quad (5.23)$$

Recall from Section 3 that the notation $\mu(t_0 + dy)$ denotes the translation of the measure μ by t_0 ; thus (5.23) means that, for a test function $h \geq 0$,

$$\int h(y) d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int_{T_{\text{fr}}^{(j)} - T_v}^{\infty} h(y - (T_{\text{fr}}^{(j)} - T_v)) d\mu_n(y). \quad (5.24)$$

Lemma 5.6. *Almost surely, for $j = 1, 2$,*

$$s_n \leq \int e^{-\lambda_n y} d\mu_{n,\text{fr}}^{(j)}(y) \leq s_n + 1. \quad (5.25)$$

Proof. By (5.24),

$$\int e^{-\lambda_n y} d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int_{T_{\text{fr}}^{(j)} - T_v}^{\infty} e^{-\lambda_n(y - (T_{\text{fr}}^{(j)} - T_v))} d\mu_n(y). \quad (5.26)$$

The expression in (5.26) is the value of the process $\sum_{v \in \text{BP}_t^{(j)}} \int_{t - T_v}^{\infty} e^{-\lambda_n(y - (t - T_v))} d\mu_n(y)$ from Definition 2.12, stopped at $t = T_{\text{fr}}^{(j)}$ (recall that $\mathcal{B}_{\text{fr}}^{(j)} = \text{BP}_{T_{\text{fr}}^{(j)}}^{(j)}$). Since μ_n has no atoms, this process is continuous in t except for jumps at the birth times, and since the birth times are distinct a.s., the corresponding jump has size $\int_0^{\infty} e^{-\lambda_n y} d\mu_n(y) = 1$. By definition, $T_{\text{fr}}^{(j)}$ is the first time the process in (5.26) exceeds s_n , so it can have value at most $s_n + 1$ at that time. \square

To prove Theorem 2.17 (a), we will separate vertices $v \in \mathcal{B}_{\text{fr}}^{(j)}$ according to whether their age at freezing, $T_{\text{fr}}^{(j)} - T_v$, is large or small. We then use Lemma 5.5 and Lemma 5.6, respectively, to bound the number of such vertices.

Proof of Theorem 2.17 (a). Lemma 5.5 and Proposition 2.11 imply that the sum in (5.18) is $O_{\mathbb{P}}(s_n)$. By (5.17), any vertex v with $T_{\text{fr}}^{(j)} - T_v \geq f_n(1) + f_n(1 + 1/s_n)$ satisfies $T_{\text{lucky}}^{(j)} - T_v \geq f_n(1 + 1/s_n)$ and therefore must contribute at least $1/s_n$ to the sum in (5.18). Consequently there can be at most $O_{\mathbb{P}}(s_n^2)$ vertices of age at least $f_n(1) + f_n(1 + 1/s_n)$ in $\mathcal{B}_{\text{fr}}^{(j)}$.

For the vertices of small age, recall the definition of $\mu_{n,\text{fr}}^{(j)}$ from (5.23) and that $\int e^{-\lambda_n y} d\mu_{n,\text{fr}}^{(j)}(y) \leq s_n + 1$ by Lemma 5.6. Apply Lemma 3.5 with K chosen large enough that $Kf_n(1) \geq f_n(1 + 1/s_n) + f_n(1)$ for each n (such a K exists since $\lim_{n \rightarrow \infty} f_n(1 + 1/s_n)/f_n(1) = e$), to see that summands corresponding to vertices of age at most $Kf_n(1)$ in $\mathcal{B}_{\text{fr}}^{(j)}$ contribute at least ε_K/s_n to the integral in Lemma 5.6. Hence, there can be at most $s_n(s_n + 1)/\varepsilon_K$ such vertices. \square

For future reference, we now state a lemma showing that most of the mass of the frozen intensity measures $\mu_{n,\text{fr}}^{(j)}$ comes from small times:

Lemma 5.7. *Let $\delta, \delta' > 0$ be given. Then there exists $K < \infty$ and $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,*

$$\mathbb{P} \left(\int e^{-\lambda_n y} \mathbb{1}_{\{\lambda_n y \geq K\}} d\mu_{n,\text{fr}}^{(j)}(y) > \delta s_n \right) \leq \delta'. \quad (5.27)$$

Proof. Let $\varepsilon = e^{-\gamma}/2$, where γ denotes Euler's constant. Using the definition of $\mu_{n,\text{fr}}^{(j)}$ from (5.23), and Lemma 2.9, we obtain $n_0 \in \mathbb{N}$ such that for all $K < \infty$, and $n \geq n_0$,

$$\int e^{-\lambda_n y} \mathbb{1}_{\{\lambda_n y \geq K\}} d\mu_{n,\text{fr}}^{(j)}(y) \leq \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int e^{-\varepsilon y / f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(T_{\text{fr}}^{(j)} - T_v + dy). \quad (5.28)$$

According to Lemma 3.4, for any $\varepsilon' > 0$ we can choose some $K < \infty$ such that, after possibly increasing n_0 , the right-hand side of (5.28) is bounded from above by $|\mathcal{B}_{\text{fr}}^{(j)}| \varepsilon' / s_n$. Since $|\mathcal{B}_{\text{fr}}^{(j)}| = O_{\mathbb{P}}(s_n^2)$ by Theorem 2.17 (a), the proof is complete. \square

5.4 Emergence of R -lucky vertices: Proof of Lemma 2.14

In this section we show how to express $\mathcal{B}_t \setminus \mathcal{B}_{\text{fr}}$, $t \geq T_{\text{unfr}}$, as a suitable union of branching processes. This representation will be useful in the proof of Lemma 2.14.

Consider the immediate children $v \in \partial \mathcal{B}_{\text{fr}}$ of individuals in the frozen cluster \mathcal{B}_{fr} . Then, for $t' \geq 0$,

$$\mathcal{B}_{T_{\text{unfr}}+t'} \setminus \mathcal{B}_{\text{fr}} = \bigcup_{v \in \partial \mathcal{B}_{\text{fr}} : T_v^{\mathcal{B}} \leq T_{\text{unfr}}+t'} \left\{ vw : w \in \text{BP}_{t'+T_{\text{unfr}}-T_v^{\mathcal{B}}}^{(v)} \right\}, \quad (5.29)$$

where $\text{BP}^{(v)}$ denotes the branching process of descendants of v , re-rooted and time-shifted as in (2.14). Furthermore, conditional on \mathcal{B}_{fr} , the children $v \in \partial \mathcal{B}_{\text{fr}}$ appear according to a Cox process. Formally, the point measures

$$\mathcal{P}_{n,\text{unfr}}^{(j)} = \sum_{v \in \partial \mathcal{B}_{\text{fr}}^{(j)}} \delta_{(T_v^{\mathcal{B}} - T_{\text{unfr}}, \text{BP}^{(v)})} \quad (5.30)$$

form Cox processes with intensities $d\mu_{n,\text{fr}}^{(j)} \otimes d\mathbb{P}(\text{BP}^{(1)} \in \cdot)$, $j = 1, 2$, where the frozen intensity measures $\mu_{n,\text{fr}}^{(j)}$ were introduced in (5.23).

Proof of Lemma 2.14. Let $\varepsilon > 0$ be given. We first claim that there exists $r > 0$ and $C < \infty$ such that $T_{v_{j,r}}^{\mathcal{B}} \leq T_{\text{unfr}} + C f_n(1)$ with probability at least $1 - \varepsilon$ for n sufficiently large. To prove this, apply Lemma 5.6 and Lemma 5.7 (with $\delta = \frac{1}{2}$, $\delta' = \varepsilon/2$) to find $K < \infty$ such that

$$\mu_{n,\text{fr}}^{(j)}(0, K/\lambda_n) \geq \int e^{-\lambda_n y} \mathbb{1}_{\{\lambda_n y \leq K\}} d\mu_{n,\text{fr}}^{(j)}(y) \geq s_n - \int e^{-\lambda_n y} \mathbb{1}_{\{\lambda_n y > K\}} d\mu_{n,\text{fr}}^{(j)}(y) \geq \frac{1}{2} s_n \quad (5.31)$$

with probability at least $1 - \frac{1}{2}\varepsilon$ for n sufficiently large. Use Lemma 2.9 to choose $C < \infty$ such that $K/\lambda_n \leq C f_n(1)$ for n large enough. Conditional on \mathcal{B}_{fr} , each vertex $v \in \partial \mathcal{B}_{\text{fr}}^{(j)}$ has an independent chance of being r -lucky, so the number of r -lucky vertices $v \in \partial \mathcal{B}_{\text{fr}}^{(j)}$ with $T_v \in (T_{\text{unfr}}, T_{\text{unfr}} + C f_n(1))$ is Poisson with mean $\mu_{n,\text{fr}}^{(j)}(0, C f_n(1)) \mathbb{P}(v \text{ is } r\text{-lucky})$. By taking r sufficiently small, Proposition 2.11 allows us to make $(\frac{1}{2} s_n) \mathbb{P}(v \text{ is } r\text{-lucky})$ large enough (uniformly in n

sufficiently large) so that, by (5.31), the number of such vertices v will be positive with probability at least $1 - \varepsilon$. This proves the claim.

It now suffices to show that $T_{v_{j,R}}^{\mathcal{B}} - T_{v_{j,r}}^{\mathcal{B}} = O_{\mathbb{P}}(f_n(1))$. The argument is very similar to the previous one, except that now we have access to at least rs_n^2 vertices whose ages are known to be small.

Let $\mathcal{L}_r^{(j)}$ denote the collection of descendants v of $v_{j,r}$ such that $T_v^{\mathcal{B}} - T_{v_{j,r}}^{\mathcal{B}} \leq f_n(1)$, so that $|\mathcal{L}_r^{(j)}| \geq rs_n^2$ by definition. Conditional on $\mathcal{L}_r^{(j)}$, for every $K' \in (0, \infty)$, the number of children w of vertices $v \in \mathcal{L}_r^{(j)}$ such that $T_w^{\mathcal{B}} - T_v^{\mathcal{B}} \in (f_n(1), f_n(1 + K'/s_n))$ is Poisson with mean at least $(rs_n^2)(K'/s_n)$. Given $\varepsilon > 0$, Proposition 2.11 allows us to choose K' large enough, so that at least one such vertex will be R -lucky with probability at least $1 - \varepsilon - o(1)$. By Condition 1.3, $f_n(1 + K'/s_n) = O(f_n(1))$, and this completes the proof. \square

6 IP and the geometry of the frozen cluster

In this section, we compare the frozen cluster to the IP cluster $\text{IP}^{(j)}(\infty)$ – the set of all vertices ever invaded in the IP process. The structure of the IP cluster is encoded in a single infinite backbone and an associated process of maximum weights, with off-backbone branches expressed in terms of Poisson Galton-Watson trees. See Proposition 6.2 below.

The proofs in this section rely on detailed comparisons between the frozen cluster and the part of the IP cluster within distance of order s_n from the root. Specifically, we show that (a) the freezing time can be effectively bounded by the time $T_{V_{[Ks_n]}^{\text{BB},j}}$ when the first passage exploration process first explores to distance Ks_n along the IP backbone (Lemma 6.6); (b) the time to complete this exploration is comparable to the largest first passage weight along the path (Lemma 6.7); and (c) the likelihood of exploring very long paths that do not belong to the IP cluster is moderate. Assertions (a) and (b) will allow us to prove Theorem 2.13, and assertion (c) will be made precise in Lemma 6.8 and Lemma 6.9 and the proof of Theorem 2.17 (b).

For the proof of Theorems 2.13 and 2.17 (b), we will obtain bounds on $\text{BP}_T^{(j)}$ for certain random times T satisfying $T \geq T_{\text{fr}}^{(j)}$ with large probability. Since $\mathcal{B}^{(j)}$ and $\text{BP}^{(j)}$ evolve in the same way until the time $T_{\text{fr}}^{(j)}$, such bounds will apply *a fortiori* to $\mathcal{B}_{\text{fr}}^{(j)}$, and we write $\text{BP}_{\text{fr}}^{(j)} = \mathcal{B}_{\text{fr}}^{(j)} = \mathcal{B}_{T_{\text{fr}}^{(j)}}^{(j)} = \text{BP}_{T_{\text{fr}}^{(j)}}^{(j)}$.

Throughout this section, we assume Conditions 1.3, 1.4 and 2.8. We will reserve the notation ε_0, δ_0 for some fixed choice of the constants in Conditions 1.4 and 2.8, with ε_0 chosen small enough to satisfy both conditions.

6.1 Structure and scaling of the IP cluster

Our description of $\text{IP}^{(j)}(\infty)$ is based on [3], which examines the structure of the IP cluster on the PWIT, and the scaling limit results in [7], which proves similar results for regular trees. As remarked in [3], the scaling limit results of [7] can be transferred to the PWIT without difficulty.

To describe the structure of the IP cluster, we first define the *backbone*:

Definition 6.1. The *backbone* of the IP cluster $\text{IP}^{(j)}(\infty)$ is the unique infinite oriented path in $\text{IP}^{(j)}(\infty)$ starting at the root. That is, the backbone is the unique (random) sequence of vertices $V_0^{\text{BB},j}, V_1^{\text{BB},j}, \dots \in \text{IP}^{(j)}(\infty)$ with $V_0^{\text{BB},j} = \emptyset_j$ and $p(V_k^{\text{BB},j}) = V_{k-1}^{\text{BB},j}$ for all $k \in \mathbb{N}$.

The PWIT edge weight between $V_{k-1}^{\text{BB},j}$ and $V_k^{\text{BB},j}$ is denoted $X_k^{\text{BB},j}$. We define the forward maximum $M_k^{(j)}$ by

$$M_k^{(j)} = \sup_{i > k} X_i^{\text{BB},j}. \quad (6.1)$$

The off-backbone branch at height k means the subtree of $\text{IP}^{(j)}(\infty)$ consisting of vertices that are descendants of $V_k^{\text{BB},j}$ but not descendants of $V_{k+1}^{\text{BB},j}$, and is denoted by τ_k . We consider τ_k as a rooted labelled tree, but with the edge weights and vertex labels from the PWIT forgotten⁵.

In the notation of [Definition 6.1](#), the maximum invaded edge weight $M^{(j)}$ from (2.5) is now $M_0^{(j)}$. (This amounts to the observation, elementary to verify, that the largest edge weight $M^{(j)}$ must occur as one of the backbone edge weights $X_k^{\text{BB},j}$.)

Proposition 6.2 ([3, 7]). *The backbone is well-defined, and $M_k^{(j)} > 1$ for each k , a.s. Furthermore:*

- (a) *The maximum in (6.1) is attained uniquely, for each k , a.s. Writing I_k for the random height at which the maximum in (6.1) is attained, it holds that $I_k = O_{\mathbb{P}}(k)$ for each $k \geq 1$.*
- (b) *The sequence $(M_k^{(j)})_{k=0}^{\infty}$ is non-increasing and forms a Markov chain with initial distribution $\mathbb{P}(M_0^{(j)} \leq m) = \theta(m)$ and transition mechanism*

$$\begin{aligned} \mathbb{P}(M_{k+1}^{(j)} = m \mid M_k^{(j)} = m) &= m(1 - \theta(m)), & m > 1, \\ \mathbb{P}(M_{k+1}^{(j)} < m' \mid M_k^{(j)} = m) &= \frac{\theta(m')}{\theta(m)}(1 - m(1 - \theta(m))), & 1 \leq m' \leq m. \end{aligned} \quad (6.2)$$

- (c) *$M_k^{(j)} = 1 + \Theta_{\mathbb{P}}(1/k)$ and indeed $k(M_k^{(j)} - 1)$ converges weakly to an exponential distribution with mean 1 as $k \rightarrow \infty$.*
- (d) *Conditional on $(M_k^{(j)})_{k=0}^{\infty}$, the off-backbone branches $(\tau_k)_{k=0}^{\infty}$ are distributed as subcritical Poisson Galton–Watson trees with means $M_k^{(j)}(1 - \theta(M_k^{(j)}))$, conditionally independent (but not identically distributed) for each k .*
- (e) *Conditional on $(M_k^{(j)})_{k=0}^{\infty}$, the PWIT edge weight $X_k^{\text{BB},j}$ either (i) equals $M_{k-1}^{(j)}$, if $M_k^{(j)} < M_{k-1}^{(j)}$; or (ii) has the Uniform $[0, M_k^{(j)}]$ distribution, if $M_k^{(j)} = M_{k-1}^{(j)}$. Furthermore the weights are conditionally independent for each k .*
- (f) *Conditional on $(M_k^{(j)})_{k=0}^{\infty}$ and $(\tau_k)_{k=0}^{\infty}$, the PWIT edge weight of an edge between two vertices of τ_k has the Uniform $[0, M_k^{(j)}]$ distribution, conditionally independent over the choice of edge and of k .*
- (g) *Conditional on $(M_k^{(j)})_{k=0}^{\infty}$ and $(\tau_k)_{k=0}^{\infty}$, the collection of PWIT edge weights between a vertex $v \in \tau_k$ and all child vertices vi for which $vi \notin \tau_k$, forms a Poisson point process of intensity 1 on the interval $(M_k^{(j)}, \infty)$. Moreover these Poisson point processes are conditionally independent for every k and every $v \in \tau_k$.*
- (h) *The part of $\text{IP}^{(j)}(\infty)$ not descended from $V_k^{\text{BB},j}$ has diameter $O_{\mathbb{P}}(k)$.*

Proof. The backbone is well-defined by Corollary 22 in [3]. The same paper proves (a) in Theorems 21 and 30, (b) in Section 3.3, (d) in Theorem 31, and (e) in Theorem 3. It has been

⁵As in the proof of [Proposition 2.11](#), we should consider instead of τ_k the set $\tilde{\tau}_k$ where we replace each vertex $v \in \tau_k \setminus \{V_k^{\text{BB},j}\}$ by an arbitrary label $\ell(v)$ drawn independently from some continuous distribution. By a slight abuse of notation, we will refer to τ and X_v , $v \in \tau_k \setminus \{V_k^{\text{BB},j}\}$ instead of $\tilde{\tau}_k$ and $X_{\ell^{-1}(v)}$, $v \in \tilde{\tau}_k$. This procedure avoids the complication, implicit in our Ulam–Harris notation, that the vertex $v = \emptyset_j k_1 k_2 \dots k_r \in \mathcal{T}^{(j)}$ automatically gives information about the number of its siblings with smaller edge weights.

observed on the top of page 954 in [3] that the methodology of [7] can be applied to show that [7, Proposition 3.3] holds for the PWIT, proving (c).

For parts (f) and (g), notice that the event that τ_k equals a particular finite tree τ requires that the children of $V_k^{\text{BB},j}$ should consist of (i) the child $V_{k+1}^{\text{BB},j}$ with edge weight consistent with the process $M^{(j)}$; and (ii) other children, and their descendants, joined to $V_k^{\text{BB},j}$ by edges of weight less than $M_k^{(j)}$, in numbers corresponding to the structure specified by τ . However, conditioning on $(M_k^{(j)})_{k=0}^\infty$ and $(\tau_k)_{k=0}^\infty$ does not impose any constraint on the precise value of the edge weights less than $M_k^{(j)}$, nor on the uninherited edge weights that exceed $M_k^{(j)}$. Parts (f) and (g) therefore follow from properties of Poisson point processes.

For (h) it suffices to notice that (d) implies that the diameter of $\bigcup_{j=0}^{k-1} \tau_j$ is stochastically dominated by k plus the maximum of k extinction times from k independent critical Poisson Galton–Watson branching processes. Since the probability that a critical Poisson Galton–Watson branching process lives to generation ℓ is $O(1/\ell)$, the claim follows. \square

We next give two lemmas that we will use in the following section to bound expectations of functions of the backbone edge weights $X_k^{\text{BB},j}$:

Lemma 6.3. *There is a constant $K < \infty$ such that, for any non-negative measurable function h and any $k \in \mathbb{N}$,*

$$\mathbb{E} \left(h(X_k^{\text{BB},j}) \mid M_0^{(j)}, \dots, M_{k-1}^{(j)} \right) \leq K(M_{k-1}^{(j)} - 1)h(M_{k-1}^{(j)}) + \int_0^1 h(M_{k-1}^{(j)}x)dx. \quad (6.3)$$

Proof. This is immediate from Proposition 6.2 (e) and the bound $\mathbb{P}(M_k^{(j)} < M_{k-1}^{(j)} \mid M_{k-1}^{(j)}) = 1 - M_{k-1}^{(j)}(1 - \theta(M_{k-1}^{(j)})) = O(M_{k-1}^{(j)} - 1)$ from (6.2) and Proposition 5.1 (b). \square

Lemma 6.4. *Given $m_0 \in (1, \infty)$, there is a constant $K < \infty$ such that, for any $k, k_0 \in \mathbb{N}_0$ with $k > k_0$ and for all $m \in (1, m_0]$, the law of $M_k^{(j)}$ conditional on $M_{k_0}^{(j)} = m$ is stochastically dominated by $m \wedge (1 + \frac{K}{k-k_0}E)$, where E is an exponential random variable with mean 1.*

Proof. Use (6.2) and (5.2) to obtain $K \in (0, \infty)$ such that, for all $k \in \mathbb{N}_0$, $1 < m' \leq m \leq m_0$,

$$\mathbb{P}(M_{k+1}^{(j)} < m' \mid M_k^{(j)} = m) = \theta(m') \frac{1 - m(1 - \theta(m))}{\theta(m)} \geq \frac{m' - 1}{K}. \quad (6.4)$$

Moreover, for $1 \leq m' \leq m \leq m_0$,

$$\mathbb{P}(m \wedge (1 + KE) < m') = 1 - e^{-(m'-1)/K} \leq \frac{m' - 1}{K}. \quad (6.5)$$

Hence, for all $k \in \mathbb{N}_0$, $1 \leq m' \leq m \leq m_0$,

$$\mathbb{P}(M_{k+1}^{(j)} < m' \mid M_k^{(j)} = m) \geq \mathbb{P}(m \wedge (1 + KE) < m') \quad (6.6)$$

and the statement of the lemma is proved for $k = k_0 + 1$. To establish the result for general $k > k_0$, we use an induction over k . According to (6.6), we can couple $M_{k+1}^{(j)}$ given $M_k^{(j)}$ to an exponential random variable E' of mean 1 that is independent of $M_k^{(j)}$ and satisfies $M_{k+1}^{(j)} \leq M_k^{(j)} \wedge (1 + KE')$ on $\{M_k^{(j)} \leq m_0\}$. Using the Markov property of $(M_k^{(j)})_k$ from Proposition 6.2 (b),

$$\mathbb{P}(M_{k+1}^{(j)} < m' \mid M_{k_0}^{(j)} = m) \geq \mathbb{E} \left(\mathbb{P}(M_k^{(j)} \wedge (1 + KE') < m' \mid M_k^{(j)}) \mid M_{k_0}^{(j)} = m \right). \quad (6.7)$$

By the induction hypothesis, the distribution of $M_k^{(j)}$ conditional on $M_{k_0}^{(j)} = m$ is stochastically dominated by $m \wedge (1 + \frac{K}{k-k_0}E'')$ for an exponential random variable E'' of mean 1 which can be chosen independent of E' . Since $x \mapsto \mathbb{P}(x \wedge (1 + KE') < m')$ is non-increasing, we obtain

$$\mathbb{P}(M_{k+1}^{(j)} < m' \mid M_{k_0}^{(j)} = m) \geq \mathbb{P}\left(m \wedge \left(1 + \frac{K}{k-k_0}E''\right) \wedge (1 + KE') < m'\right). \quad (6.8)$$

Since E' and E'' are independent, $(\frac{1}{k-k_0}E'') \wedge E'$ is equal in distribution to $\frac{1}{k+1-k_0}E$ for an exponential random variable E of mean 1 and the proof is complete. \square

6.2 First passage times and the IP cluster: Proof of Theorem 2.13

For notational convenience, given a constant $K < \infty$ to be fixed below, we will abbreviate

$$T = T_{V_{\lfloor Ks_n \rfloor}^{\text{BB},j}}. \quad (6.9)$$

for the remainder of this section. We begin with the following preliminary lemma:

Lemma 6.5. *Let $(U_i)_{i \in \mathbb{N}}$ be an i.i.d. sequence of uniform random variables on $(0, 1)$. There exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that*

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) \leq f_n(m)\right) \geq \delta \quad \text{for all } m \in [1, \infty), n \geq n_0. \quad (6.10)$$

Proof. Choose $u \in (0, 1)$ such that $u^{\varepsilon_0} \leq \frac{1}{4}\varepsilon_0$, where ε_0 was fixed below Condition 2.8. Using the inequality $\mathbb{P}(U_i \leq u^{1/s_n} \forall i \leq \lfloor s_n \rfloor) \geq u$, we can bound $\mathbb{P}(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) \leq f_n(m)) \geq u\mathbb{P}\left(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) \leq f_n(m) \mid U_i \leq u^{1/s_n} \forall i \leq \lfloor s_n \rfloor\right)$. By Lemma 5.3, for sufficiently large n and for all $m \geq 1$,

$$\begin{aligned} & \mathbb{E}\left(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) \mid U_i \leq u^{1/s_n} \forall i \leq \lfloor s_n \rfloor\right) \\ & \leq s_n \int_0^1 f_n(u^{1/s_n}mx)dx \leq f_n(m)\left(\frac{u^{\varepsilon_0}}{\varepsilon_0} + \frac{f_n(1)}{f_n(m)} \frac{(1-\delta_0)^{\varepsilon_0 s_n} s_n}{u^{1/s_n} m}\right) \leq (\tfrac{1}{4} + o(1))f_n(m). \end{aligned} \quad (6.11)$$

By Markov's inequality, $\mathbb{P}\left(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) > f_n(m) \mid U_i \leq u^{1/s_n} \forall i \leq \lfloor s_n \rfloor\right) \leq (\tfrac{1}{4} + o(1)) \leq \tfrac{1}{2}$, so that $\mathbb{P}\left(\sum_{i=1}^{\lfloor s_n \rfloor} f_n(mU_i) \leq f_n(m) \mid U_i \leq u^{1/s_n} \forall i \leq \lfloor s_n \rfloor\right) \geq \tfrac{1}{2}$ for all n sufficiently large. We may therefore take $\delta = u/2$. \square

Lemma 6.6. *Given $\delta > 0$, there exist $K < \infty$ and $n_0 \in \mathbb{N}$ such that $T_{\text{fr}}^{(j)} \leq T_{V_{\lfloor Ks_n \rfloor}^{\text{BB},j}}$ with probability at least $1 - \delta$ for all $n \geq n_0$.*

Proof. By Theorem 2.17 (a) there exists K_1 such that $\mathbb{P}(|\text{BP}_{\text{fr}}^{(j)}| > K_1 s_n^2) < \delta/4$ for large n . Recall from (6.9) that we write $T = T_{V_{\lfloor Ks_n \rfloor}^{\text{BB},j}}$. We now show that there is a $K < \infty$ such that $\text{BP}_T^{(j)}$ contains more than $K_1 s_n^2$ vertices with probability at least $1 - 3\delta/4$ for large n . This will then prove the lemma. Set

$$\mathcal{A} = \{M_{\lfloor s_n \rfloor}^{(j)} \leq 1 + K_2/s_n\}, \quad (6.12)$$

where $K_2 < \infty$ is chosen using [Proposition 6.2 \(c\)](#) so that $\mathbb{P}(\mathcal{A}) > 1 - \delta/4$ for large n . By monotonicity, on \mathcal{A} , $M_k^{(j)} \leq 1 + K_2/s_n$ for all $k \geq \lfloor s_n \rfloor$.

Let $k \geq \lfloor s_n \rfloor$ and write $\tau'_k \subset \tau_k$ for those vertices in the k^{th} off-backbone branch that lie within distance s_n of $V_k^{\text{BB},j}$. By [Proposition 6.2 \(d\)](#), given $M_k^{(j)}$, τ_k is a Poisson Galton–Watson tree with mean $\hat{M}_k^{(j)} = M_k^{(j)}(1 - \theta(M_k^{(j)}))$. When $M_k^{(j)} \in (1, 1 + K_2/s_n]$, then [\(5.3\)](#) yields K_3 such that $\hat{M}_k^{(j)} \in [1 - K_3/s_n, 1)$ for sufficiently large n . Hence, [Proposition 5.1 \(e\)](#) and [\(f\)](#) yield $\delta_1 > 0$ such that $\mathbb{P}(|\tau'_k| \geq s_n^2 \mid M_k^{(j)} \leq 1 + K_1/s_n) \geq \delta_1/s_n$ for sufficiently large n .

Consider $v \in \tau'_k$. By [Proposition 6.2 \(f\)](#), conditional on $M_k^{(j)}$, the path from $V_k^{\text{BB},j}$ to v consists of at most s_n edges whose PWIT edge weights X_{w_j} are independent and uniformly distributed on $[0, M_k^{(j)}]$. Write τ''_k for the collection of vertices $v \in \tau'_k$ such that $T_v - T_{V_k^{\text{BB},j}} \leq f_n(M_k^{(j)})$. By [Lemma 6.5](#), there exists $\delta_2 > 0$ such that $\mathbb{P}(v \in \tau''_k \mid v \in \tau'_k, M_k^{(j)}) \geq \delta_2$, so that [Lemma 5.2](#) implies that $\mathbb{P}(|\tau''_k| \geq \frac{1}{2}\delta_2 s_n^2 \mid M_k^{(j)}) \geq (\frac{1}{2}\delta_2)(\delta_1/s_n)$ on \mathcal{A} .

By [Proposition 6.2 \(d\)](#) and [\(f\)](#), the off-backbone branches $\tau_k, \tau'_k, \tau''_k$ are conditionally independent for different k given $(M_k^{(j)})_{k=0}^\infty$. Therefore, conditional on $(M_k^{(j)})_{k=0}^\infty$ and the event \mathcal{A} , we have for each $k \geq \lfloor s_n \rfloor$ an independent chance, bounded below by $\frac{1}{2}\delta_2\delta_1/s_n$, that $|\tau''_k| \geq \frac{1}{2}\delta_2 s_n^2$. It follows that, by taking K_2 sufficiently large, we have $\mathbb{P}\left(\sum_{k=\lfloor s_n \rfloor+1}^{\lfloor K_2 s_n \rfloor} |\tau''_k| > K_1 s_n^2 \mid \mathcal{A}\right) > 1 - \delta/4$.

Finally, let \mathcal{A}' be the event that the unique edge with weight $M_{\lfloor K_2 s_n \rfloor}^{(j)}$ occurs at height at most $K s_n$. By [Proposition 6.2 \(a\)](#), we may choose $K > K_2$ large enough that $\mathbb{P}(\mathcal{A}') > 1 - \delta/4$.

Whenever \mathcal{A}' occurs, it follows that $T - T_{V_k^{\text{BB},j}} \geq f_n(M_k^{(j)})$ for each $k \leq K_2 s_n$, since the collection of first passage weights $f_n(X_{k+1}^{\text{BB},j}), \dots, f_n(X_{\lfloor K s_n \rfloor}^{\text{BB},j})$ along the path from $V_k^{\text{BB},j}$ to $V_{\lfloor K s_n \rfloor}^{\text{BB},j}$ must contain the edge weight $f_n(M_k^{(j)})$. In particular, for any $v \in \tau''_k$ with $k \leq K_2 s_n$, we have $T_v \leq f_n(M_k^{(j)}) + T_{V_k^{\text{BB},j}} \leq T$ and therefore $v \in \text{BP}_T^{(j)}$. Thus

$$\begin{aligned} \mathbb{P}(T < T_{\text{fr}}^{(j)}) &\leq \mathbb{P}(|\text{BP}_{\text{fr}}^{(j)}| > K_1 s_n^2) + \mathbb{P}(T < T_{\text{fr}}^{(j)} \text{ and } |\text{BP}_{\text{fr}}^{(j)}| \leq K_1 s_n^2) \\ &< \delta/4 + \mathbb{P}(|\text{BP}_T^{(j)}| \leq K_1 s_n^2) \\ &\leq \delta/4 + \mathbb{P}(\mathcal{A}^c) + \mathbb{P}((\mathcal{A}')^c) + \mathbb{P}(\mathcal{A}' \cap \{|\text{BP}_T^{(j)}| \leq K_1 s_n^2\} \mid \mathcal{A}) \\ &< \delta/4 + \delta/4 + \delta/4 + \mathbb{P}\left(\sum_{k=\lfloor s_n \rfloor+1}^{\lfloor K_2 s_n \rfloor} |\tau''_k| \leq K_1 s_n^2 \mid \mathcal{A}\right) < \delta. \end{aligned} \quad \square$$

Lemma 6.7. *Given $m_0 \in (1, \infty)$ and $K < \infty$, there is a constant $K'' < \infty$ such that*

$$\mathbb{E}\left(T_{V_{\lfloor K s_n \rfloor}^{\text{BB},j}} - T_{V_{k_0}^{\text{BB},j}} \mid M_{k_0}^{(j)} = m\right) \leq K'' f_n(m) \quad (6.13)$$

for all $k_0 \leq \lfloor K s_n \rfloor$ and all $1 \leq m \leq m_0$.

Proof. We have $T_{V_{\lfloor K s_n \rfloor}^{\text{BB},j}} - T_{V_{k_0}^{\text{BB},j}} = \sum_{k=k_0+1}^{\lfloor K s_n \rfloor} f_n(X_k^{\text{BB},j})$. Using first [Lemma 6.3](#) to find a constant

$K_1 < \infty$ and then [Lemma 5.3](#) and [Lemma 3.1](#) and $M_{k-1}^{(j)} \leq M_{k_0}^{(j)}$, we estimate uniformly in m

$$\begin{aligned}
& \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(f_n(X_k^{\text{BB},j}) \mid M_{k_0}^{(j)} = m \right) \\
& \leq \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(K_1 (M_{k-1}^{(j)} - 1) f_n(M_{k-1}^{(j)}) + \int_0^1 f_n(M_{k-1}^{(j)} x) dx \mid M_{k_0}^{(j)} = m \right) \\
& \leq \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(K_1 (M_{k-1}^{(j)} - 1) f_n(M_{k-1}^{(j)}) + \frac{f_n(M_{k-1}^{(j)}) + o(f_n(1))}{\varepsilon_0 s_n} \mid M_{k_0}^{(j)} = m \right) \\
& \leq O(f_n(m)) \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left((M_{k-1}^{(j)} - 1) \left(\frac{M_{k-1}^{(j)}}{m} \right)^{\varepsilon_0 s_n} \mid M_{k_0}^{(j)} = m \right) + O(f_n(m)) + o(f_n(1)). \quad (6.14)
\end{aligned}$$

In the sum on the right-hand side of (6.14), the term $k = k_0 + 1$ contributes at most $m_0 - 1$ since $m \leq m_0$. The sum of contributions from the events $\{M_{k-1}^{(j)} \leq 1 + 1/s_n\}$, $k = k_0 + 2, \dots, \lfloor Ks_n \rfloor$, is $O(1)$ since $M_{k-1}^{(j)} \leq M_{k_0}^{(j)}$. We conclude that

$$\begin{aligned}
& \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(f_n(X_k^{\text{BB},j}) \mid M_{k_0}^{(j)} = m \right) \\
& \leq O(f_n(m)) + O(f_n(m)) \sum_{k=k_0+2}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(\mathbb{1}_{\{M_{k-1}^{(j)} \geq 1 + 1/s_n\}} (M_{k-1}^{(j)} - 1) \left(\frac{M_{k-1}^{(j)}}{m} \right)^{\varepsilon_0 s_n} \mid M_{k_0}^{(j)} = m \right). \quad (6.15)
\end{aligned}$$

In the expectation on the right-hand side of (6.15), the integrand is increasing in $M_{k-1}^{(j)}$, so as an upper bound we can use [Lemma 6.4](#) to replace $M_{k-1}^{(j)}$ by $m \wedge (1 + \frac{K'}{k-1-k_0} E)$ for some $K' < \infty$. Letting $i = k - 1 - k_0$,

$$\begin{aligned}
& \sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(f_n(X_k^{\text{BB},j}) \mid M_{k_0}^{(j)} = m \right) - O(f_n(m)) \\
& \leq O(f_n(m)) \sum_{i=1}^{\infty} \left[(m-1) \mathbb{P} \left(\frac{K'E}{i} > m-1 \right) + \mathbb{E} \left(\mathbb{1}_{\{1/s_n \leq \frac{K'E}{i} \leq m-1\}} \frac{K'E}{i} \left(\frac{1 + K'E/i}{m} \right)^{\varepsilon_0 s_n} \right) \right] \\
& \leq O(f_n(m)) \sum_{i=1}^{\infty} \left((m-1) e^{-(m-1)i/K'} + \int_{1/s_n}^{m-1} y \left(\frac{1+y}{m} \right)^{\varepsilon_0 s_n} \frac{i}{K'} e^{-iy/K'} dy \right) \\
& \leq O(f_n(m)) \left(\frac{m-1}{1 - e^{-(m-1)/K'}} + \int_{1/s_n}^{m-1} \frac{y/K'}{(1 - e^{-y/K'})^2} \left(\frac{1+y}{m} \right)^{\varepsilon_0 s_n} dy \right). \quad (6.16)
\end{aligned}$$

The ratio $y/(1 - e^{-y/K'})$ is uniformly bounded for $y \leq m_0 - 1$ and $1/(1 - e^{-y/K'}) \leq 1/(1 - e^{-1/(K's_n)}) \sim K's_n$ for $y \geq 1/s_n$. We conclude that

$$\sum_{k=k_0+1}^{\lfloor Ks_n \rfloor} \mathbb{E} \left(f_n(X_k^{\text{BB},j}) \mid M_{k_0}^{(j)} = m \right) \leq O(f_n(m)) + O(f_n(m)) \int_{1/s_n}^{m-1} s_n \left(\frac{1+y}{m} \right)^{\varepsilon_0 s_n} dy, \quad (6.17)$$

and this upper bound is $O(f_n(m))$ as required. \square

Lemma 6.6 and Lemma 6.7 now allow us to prove Theorem 2.13:

Proof of Theorem 2.13. We need to show that $f_n^{-1}(T_{\text{fr}}^{(j)}) \xrightarrow{\mathbb{P}} M_0^{(j)}$. We begin by arguing that $f_n^{-1}(T_{\text{fr}}^{(j)}) \geq M_0^{(j)}$ whp. Indeed, the subgraph $\text{BP}_{f_n(M_0^{(j)})-}^{(j)}$ explored strictly before time $f_n(M_0^{(j)})$ is a subgraph of those vertices connected to the root by paths that use only edges of PWIT edge weight strictly less than $M_0^{(j)}$. This latter subgraph is a random subgraph of $\mathcal{T}^{(j)}$ that is finite a.s. (by Proposition 6.2 (a) and (d)) and independent of n . In particular, the subgraph $\text{BP}_{f_n(M_0^{(j)})-}^{(j)}$ has size $O_{\mathbb{P}}(1)$. To show that $T_{\text{fr}}^{(j)} \geq f_n(M_0^{(j)})$ whp, it therefore suffices to show that $|\mathcal{B}_{\text{fr}}^{(j)}| \xrightarrow{\mathbb{P}} \infty$.

Indeed, we shall show that $\mathcal{B}_{\text{fr}}^{(j)}$ must contain at least of order s_n vertices. To this end, it suffices by Lemma 5.6 to show that any vertex $v \in \mathcal{B}_{\text{fr}}^{(j)}$ can contribute at most $O(1)$ to $\int e^{-\lambda_n y} d\mu_{n,\text{fr}}^{(j)}(y)$. We use Lemma 3.3 to compute

$$\int e^{-\lambda_n y} \mu_n(t - T_v + dy) \leq 1 + \int_{f_n(1)}^{\infty} e^{-\lambda_n y} \frac{dy}{\varepsilon_0 s_n f_n(1)} \quad (6.18)$$

for n large enough, uniformly in the age $t - T_v$, and the upper bound in (6.18) is $O(1)$ by Lemma 2.9. We have therefore shown that $f_n^{-1}(T_{\text{fr}}^{(j)}) \geq M_0^{(j)}$ whp.

For the corresponding upper bound, let $\varepsilon > 0$ be given and choose $m_0 < \infty$ large enough that $\mathbb{P}(M_0^{(j)} > m_0) < \varepsilon$. By Lemma 6.6 it is enough to show that for any fixed $K < \infty$, $T_{V_{\lfloor K s_n \rfloor}^{\text{BB},j}} \leq f_n(M_0^{(j)} + \varepsilon)$ whp on $\{M_0^{(j)} \leq m_0\}$. Use Lemma 6.7 with $k_0 = 0$ together with Markov's inequality to find a constant $K' < \infty$ such that

$$\mathbb{P}\left(T_{V_{\lfloor K s_n \rfloor}^{\text{BB},j}} \geq f_n(M_0^{(j)} + \varepsilon), M_0^{(j)} \leq m_0\right) \leq K' \mathbb{E}\left(\frac{f_n(M_0^{(j)})}{f_n(M_0^{(j)} + \varepsilon)} \mathbb{1}_{\{M_0^{(j)} \leq m_0\}}\right). \quad (6.19)$$

By Lemma 3.1, $f_n(M_0^{(j)} + \varepsilon)/f_n(M_0^{(j)}) \xrightarrow{a.s.} \infty$, and the dominated convergence theorem completes the proof. \square

6.3 Diameter of the frozen cluster: Proof of Theorem 2.17 (b)

We next give an upper bound on the number of boundary edges of the IP cluster for which a distant descendant is explored before the freezing time. Here we say that a vertex v is explored by time t if $v \in \text{BP}_t$. Vertices that belong to the IP cluster $\text{IP}(\infty)$ are called *invaded*. Write $N(k, K)$ for the number of vertices of $\partial\tau_k$ for which some descendant at least $\lceil s_n \rceil$ generations away belongs to $\text{BP}_T^{(j)}$. (Recall from (6.9) the abbreviation $T = T_{V_{\lfloor K s_n \rfloor}^{\text{BB},j}}$, and recall also that τ_k denotes the tree consisting of vertices that are descendants of $V_k^{\text{BB},j}$ but not descendants of $V_{k+1}^{\text{BB},j}$.)

Lemma 6.8. *Given $\delta > 0$, $K < \infty$ and $m_0 \in (1, \infty)$, there is a constant $K' < \infty$ such that for all $k \in \mathbb{N}_0$*

$$\mathbb{E}\left(N(k, K) \mid 1 + \delta/s_n \leq M_k^{(j)} \leq m_0\right) \leq K'/s_n. \quad (6.20)$$

Proof. Consider $v \in \tau_k$ and its children, which we write as vi , $i \in \mathbb{N}$. Recall from Proposition 6.2 (g) that, conditional on the invasion cluster, the PWIT weights $\{X_{vi} : vi \notin \tau_k\}$ (i.e., the weights corresponding to children that were *not* invaded) form a Poisson point process on $(M_k^{(j)}, \infty)$ with intensity 1.

A child vi of v will be explored by the FPP process by time T if and only if $f_n(X_{vi}) \leq T - T_v$. Abbreviate $\tilde{X} = f_n^{-1}(T - T_{V_k^{\text{BB},j}})$. Since $T_v \geq T_{V_k^{\text{BB},j}}$, we have as an upper bound that vi can only

be explored if $X_{vi} \leq \tilde{X}$. In particular, the conditional expectation of the number of uninvaded children of v that are explored by the FPP process by time T obeys the bound

$$\mathbb{E} \left(\left| \left\{ i \in \mathbb{N} : vi \notin \tau_k, vi \in \mathbf{BP}_T^{(j)} \right\} \right| \middle| M_k^{(j)} = m, \tilde{X} \right) \leq (\tilde{X} - m) \mathbb{1}_{\{\tilde{X} \geq m\}}. \quad (6.21)$$

For each such child vi , we may bound the probability that a descendant of vi at least $\lceil s_n \rceil$ generations away is explored by time T by the probability that a Poisson Galton–Watson branching process with mean \tilde{X} survives for at least s_n generations. Because of (6.21) and since in (6.20) $M_k^{(j)} \geq 1 + \delta/s_n$, we may assume that $\tilde{X} \geq m \geq 1 + \delta/s_n$. By Proposition 5.1 (g) and (b), this probability is $O(\tilde{X} - 1)$.

On the other hand, conditional on $M_k^{(j)} = m$, the expected size of τ_k is $O(1/(m-1))$, since, by Proposition 6.2 (d) and (5.3), τ_k is distributed as a Poisson Galton–Watson tree with subcritical parameter $\hat{m} < 1$ and $1 - \hat{m} \sim m - 1$. We conclude that uniformly in $m \in [1 + \delta/s_n, \infty)$,

$$\begin{aligned} \mathbb{E} \left(N(k, K) \middle| M_k^{(j)} = m, \tilde{X} \right) &\leq O(1) \frac{(\tilde{X} - 1)(\tilde{X} - m)}{m - 1} \mathbb{1}_{\{\tilde{X} \geq m\}} \\ &= O(1) \left(\tilde{X} - m + \frac{(\tilde{X} - m)^2}{m - 1} \right) \mathbb{1}_{\{\tilde{X} \geq m\}}. \end{aligned} \quad (6.22)$$

In order to bound the right-hand side of (6.22), we first use Lemma 3.1 and Lemma 6.7 and Markov's inequality to obtain a constant $K' < \infty$ such that for all $m \in [1, m_0]$ and $x \geq 0$,

$$\begin{aligned} \mathbb{P} \left(\tilde{X} - m \geq x \middle| M_k^{(j)} = m \right) &= \mathbb{P} \left(T - T_{V_k^{\text{BB},j}} \geq f_n(m + x) \middle| M_k^{(j)} = m \right) \\ &\leq \mathbb{P} \left(T - T_{V_k^{\text{BB},j}} \geq \left(1 + \frac{x}{m}\right)^{\varepsilon_0 s_n} f_n(m) \middle| M_k^{(j)} = m \right) \leq \frac{K'}{(1 + x/m)^{\varepsilon_0 s_n}}. \end{aligned} \quad (6.23)$$

Thus, for $1 \leq m \leq m_0$ and $p \in \{1, 2\}$,

$$\begin{aligned} \mathbb{E} \left((\tilde{X} - m)^p \mathbb{1}_{\{\tilde{X} \geq m\}} \middle| M_k^{(j)} = m \right) &= \int_0^\infty p x^{p-1} \mathbb{P} \left(\tilde{X} - m \geq x \middle| M_k^{(j)} = m \right) dx \\ &\leq K' \int_0^\infty \frac{p x^{p-1} dx}{(1 + x/m)^{\varepsilon_0 s_n}} = O(1) \int_1^\infty \frac{(t-1)^{p-1} dt}{t^{\varepsilon_0 s_n}} = \begin{cases} O(1/s_n), & p = 1, \\ O(1/s_n^2), & p = 2. \end{cases} \end{aligned} \quad (6.24)$$

For $m \geq 1 + \delta/s_n$, we have $1/(m-1) \leq O(s_n)$ and we obtain that uniformly for $m \in [1 + \delta/s_n, m_0]$,

$$\mathbb{E} \left(\left(\tilde{X} - m + \frac{(\tilde{X} - m)^2}{m - 1} \right) \mathbb{1}_{\{\tilde{X} \geq m\}} \middle| M_k^{(j)} = m \right) = O(1/s_n) + O(s_n)O(1/s_n^2) = O(1/s_n). \quad (6.25)$$

Combining (6.22) with (6.25) completes the proof. \square

Lemma 6.8 will allow us to control the effect of branches τ_k for small k . For larger k , the following estimate will allow us to bound the probability that a very long path is explored between time $T_{V_k^{\text{BB},j}}$ and the freezing time. Let $v \in \tau_k$. Write $D_{v,r,C,\ell}$ for the collection of descendants of v connected to v by exactly r edges of weight at most $1 + C/s_n$, at most ℓ of which lie between $1 - 1/s_n$ and $1 + C/s_n$.

Lemma 6.9. *Given $C, \ell < \infty$, there is $C' < \infty$ such that $\mathbb{P}(D_{v,r,C,\ell} \neq \emptyset) \leq C'/r$ for all $r \in \mathbb{N}$, $v \in \mathcal{T}$ and $s_n \geq 1$.*

Proof. On the event $\{D_{v,r,C,\ell} \neq \emptyset\}$, let W be chosen uniformly from $D_{v,r,C,\ell}$, and let L denote the collection of vertices wi along the path from v to W for which $1 - 1/s_n \leq X_{wi} \leq 1 + C/s_n$. By the Poisson point process property, conditional on the occurrence of $\{D_{v,r,C,\ell} \neq \emptyset\}$ and the values of W and L , the edge weights X_{wi} , $wi \in L$, are uniformly distributed on $[1 - 1/s_n, 1 + C/s_n]$. In particular, $1 - 1/s_n \leq X_{wi} \leq 1$ for each $wi \in L$ with conditional probability $(1/(C+1))^{|L|}$, which is at least $(1/(C+1))^\ell$ by construction. If this occurs then v is connected by edges of weight at most 1 to a descendant at generation r . Using the notation from [Proposition 5.1](#) and appealing to standard properties of critical Galton–Watson processes (see for example Lemma I.1.10 in [\[15\]](#)),

$$O(1/r) = \mathbb{P}_1(|\tau^{(\geq r)}| \geq 1) \geq \mathbb{P}(D_{v,r,C,\ell} \neq \emptyset) \cdot (1/(C+1))^\ell, \quad (6.26)$$

completing the proof. \square

Proof of Theorem 2.17 (b). We have to show that the diameter of $\mathcal{B}_{\text{fr}}^{(j)}$ is $O_{\mathbb{P}}(s_n)$. Let $\varepsilon > 0$ be given. Use [Lemma 6.6](#) to choose $K < \infty$ such that $T_{\text{fr}}^{(j)} \leq T$ with probability at least $1 - \varepsilon$ for large n . Use [Proposition 6.2 \(c\)](#) to choose $\delta > 0$ and $m_0 < \infty$ so that $1 + \delta/s_n \leq M_{\lfloor Ks_n \rfloor}^{(j)} \leq M_0^{(j)} \leq m_0$ with probability at least $1 - \varepsilon$. By [Lemma 6.8](#), we may choose $\eta \in (0, 1)$ sufficiently small that $\mathbb{E} \left(\mathbb{1}_{\{1 + \delta/s_n \leq M_{\lfloor s_n \rfloor}^{(j)} \leq M_0^{(j)} \leq m_0\}} \sum_{k=0}^{\lfloor \eta s_n \rfloor} N(k, K) \right) < \varepsilon$, so that $N(k, K) = 0$ for each $k \leq \eta s_n$ with probability at least $1 - 2\varepsilon$ for large n .

Use [Proposition 6.2 \(c\)](#) to choose $C' < \infty$ such that $M_{\lfloor \eta s_n \rfloor}^{(j)} \leq 1 + C'/s_n$ with probability at least $1 - \varepsilon$. Assuming that this event occurs, [Lemma 6.7](#) and Markov's inequality give $C'' < \infty$ such that $T - T_{V_{\lfloor \eta s_n \rfloor}^{\text{BB},j}} \leq C'' f_n(1 + C'/s_n)$ with conditional probability at least $1 - \varepsilon$. By [Condition 1.3](#), it follows that we may choose $C, \ell < \infty$ such that

$$T - T_{V_{\lfloor \eta s_n \rfloor}^{\text{BB},j}} \leq f_n(1 + C/s_n) \leq \ell f_n(1 - 1/s_n) \quad (6.27)$$

with probability at least $1 - \varepsilon$ for all sufficiently large n .

With these preliminaries, we can now complete the proof. Assume that all the above events occur, making an error of at most 6ε . By [Proposition 6.2 \(h\)](#), the part of the IP cluster that does not descend from $V_{\lfloor Ks_n \rfloor}^{\text{BB},j}$ has diameter $O_{\mathbb{P}}(s_n)$. Hence it suffices to show that vertices of $\text{BP}_T^{(j)}$ are within distance $O_{\mathbb{P}}(s_n)$ from the IP cluster. The assumption $N(k, K) = 0$, $k \leq \eta s_n$, verifies this for the beginning of the backbone.

For $k \geq \eta s_n$, (6.21), (6.27) and the bound $\mathbb{E}(|\tau_k| | M_k^{(j)}) = O(1/(M_k^{(j)} - 1)) = O(s_n)$ imply that the (conditionally) expected number of vertices $vi \in \text{BP}_T^{(j)} \setminus \tau_k$ for which $v \in \tau_k$, for some $\eta s_n \leq k \leq Ks_n$, is at most $(C/s_n) \cdot O(s_n) \cdot (Ks_n) = O(s_n)$. If such a vertex vi has a descendant $w \in \text{BP}_T^{(j)}$ at distance r , then by (6.27) the path from vi to w can contain no edges of PWIT weight greater than $1 + C/s_n$ and at most ℓ edges of PWIT weight greater than $1 - 1/s_n$. In other words, necessarily $D_{vi,r,C,\ell} \neq \emptyset$. This event is conditionally independent of the possible vi , so by [Lemma 6.9](#) the conditional probability of any $D_{vi,r,C,\ell}$ occurring is at most $O(s_n/r)$. This can be made smaller than ε by taking $r = C''' s_n$ for C''' large enough, and this completes the proof. \square

7 A strong disorder result: Proof of Theorem 1.6

Proof of Theorem 1.6. The proof of (1.14) proceeds via stochastic upper and lower bounds on W_n based on couplings with IP where we use two different exploration processes. For the lower bound, consider the minimal-rule exploration process $\mathcal{E} = (\mathcal{E}_k)_{k \in \mathbb{N}_0}$ given by IP on \mathcal{T} that alternates

between an invasion step in $\mathcal{T}^{(1)}$ and an invasion step in $\mathcal{T}^{(2)}$ as explained below [Definition 4.5](#). By [Theorem 4.4](#) the edge-weights $Y_e^{(K_n)} = f_n(X_e^{(K_n)})$ derived from this exploration process as in (4.4) are i.i.d. with distribution function F_Y and it suffices to consider the FPP problem on K_n with these edge weights. Let $\mathcal{V}^{(j)}$ denote the set of vertices in the invasion cluster on $\mathcal{T}^{(j)}$ explored before the edge of weight $M_0^{(j)}$ is invaded. (In the language of [\[3, 10, 14, 17\]](#), $\mathcal{V}^{(j)}$ is the first pond, not including the first outlet.) Let $\mathcal{V}_+^{(j)}$ consist of $\mathcal{V}^{(j)}$ together with all adjacent vertices connected by an edge of weight at most $M_0^{(1)} \vee M_0^{(2)}$.

Let \mathcal{A}_n be the event that none of the vertices in $\mathcal{V}_+^{(1)} \cup \mathcal{V}_+^{(2)}$ are thinned, and that the exponential variable $X_{\{1,2\}}^{(K_n)} = E_{\{1,2\}}$ from (4.4) satisfies $X_{\{1,2\}}^{(K_n)} \geq \frac{1}{n}(M_0^{(1)} \vee M_0^{(2)})$. Since $\mathcal{V}_+^{(j)}$ is finite and the vertices in $\mathcal{V}^{(j)}$ are explored after a finite number of steps (not depending on n), \mathcal{A}_n holds with high probability.

On \mathcal{A}_n , let $\mathcal{W}^{(j)}$ denote the image in $[n]$ of $\mathcal{V}^{(j)}$ under the mapping $\pi_M: v \mapsto M_v$ from [Definition 2.5](#). Then every edge e between $\mathcal{W}^{(j)}$ and $[n] \setminus (\mathcal{W}^{(1)} \cup \mathcal{W}^{(2)})$ satisfies $Y_e^{(K_n)} \geq f_n(M_0^{(j)})$, and every edge e between $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ satisfies $Y_e^{(K_n)} \geq f_n(M_0^{(1)} \vee M_0^{(2)})$. Since every path between vertices 1 and 2 has to leave $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$, this therefore proves that $W_n \geq f_n(M_0^{(1)} \vee M_0^{(2)})$ on \mathcal{A}_n , i.e., whp.

For the upper bound, let $\varepsilon > 0$ and let $N, N_1 \in \mathbb{N}$ denote constants to be chosen later. Modify the minimal-rule exploration process above by stopping after N steps in each subtree $\mathcal{T}^{(j)}$, i.e., set $\mathcal{E}'_k = \mathcal{E}_{k \wedge 2N}$, and couple the edge weights according to (4.4). Denote by $X_{\max}^{(j)}$ the largest edge weight in $\mathcal{T}^{(j)}$ so invaded, so that $X_{\max}^{(j)} \leq M_0^{(j)}$ by definition.

Let $\mathcal{U}^{(j)} = \{v \in \partial \mathcal{E}_{2N_1} \cap \mathcal{T}^{(j)} : X_{\max}^{(j)} < X_v < X_{\max}^{(j)} + \varepsilon\}$ denote the collection of boundary vertices joined to invaded vertices by an edge of weight at most $X_{\max}^{(j)} + \varepsilon$. Conditional on $X_{\max}^{(1)}, X_{\max}^{(2)}$ and \mathcal{E}' , the number $|\mathcal{U}^{(j)}|$ of such boundary vertices is Poisson with mean εN , independently for $j \in \{1, 2\}$. (This holds because the event that the exploration process \mathcal{E}' explores a given sequence of vertices v_1, \dots, v_{2N_1} can be expressed solely in terms of the numbers $|\{vw \in \partial \mathcal{E}'_k : X_{vw} < X_{v_i}\}|$ of boundary edges of *smaller* weight, over all $k, i = 1, \dots, 2N$.)

Let \mathcal{A}'_n denote the event that none of the vertices in $\mathcal{E}'_{2N} \cup \mathcal{U}^{(1)} \cup \mathcal{U}^{(2)}$ have the same mark. Condition on the occurrence of \mathcal{A}'_n and the disjoint vertex sets $\pi_M(\mathcal{E}'_{2N}), \pi_M(\mathcal{U}^{(1)}), \pi_M(\mathcal{U}^{(2)})$. Consider the induced subgraph K'_{n-2N-2} of K_n obtained by excluding the $2N + 2$ explored vertices in $\pi_M(\mathcal{E}'_{2N})$. Since no other vertices are explored, the edge weights in this induced subgraph are the independent exponential random variables E_e from (4.4).

The random subgraph $G'_{n-2N-2} = \{e \in E(K'_{n-2N-2}) : E_e \leq \frac{1}{n}(1 + \varepsilon)\}$ has the (conditional) law of the Erdős-Rényi random graph $G(n - 2N - 2, p)$ with $p = \mathbb{P}(E_e \leq \frac{1}{n}(1 + \varepsilon)) \sim \frac{1}{n}(1 + \varepsilon)$ as $n \rightarrow \infty$. As is well known, in the supercritical regime, the giant component has diameter $O_{\mathbb{P}}(\log n)$ and contains a positive asymptotic fraction of vertices (see e.g., [\[?\]](#)). Suppose $U^{(1)}, U^{(2)}$ are two disjoint subsets of vertices in G'_{n-2N-2} (possibly random but independent of the randomness in G'_{n-2N-2}). If $U^{(1)}$ and $U^{(2)}$ are sufficiently large, each of them is likely to contain at least one vertex from the giant component. Hence we may choose $N_1 \in \mathbb{N}$ such that, given the event $\{|U^{(1)}|, |U^{(2)}| \geq N_1\}$, there will exist a pair of vertices $u_1 \in U^{(1)}, u_2 \in U^{(2)}$ connected by a path in G'_{n-2N-2} of length at most $N_1 \log n$, with conditional probability at least $1 - \varepsilon$ for n sufficiently large.

Since the sizes $|\mathcal{U}^{(j)}|$ are independent Poisson random variables with mean εN , we may choose N large enough that $|\mathcal{U}^{(j)}| \geq N_1$ with probability at least $1 - \varepsilon$. Moreover, since $\mathcal{E}'_{2N}, \mathcal{U}^{(1)}, \mathcal{U}^{(2)}$ are finite and do not depend on n , it follows that \mathcal{A}_n occurs with high probability and we can choose N_2 large enough that the diameters of $\mathcal{E}'_{2N}, \mathcal{U}^{(1)}, \mathcal{U}^{(2)}$ are at most N_2 with probability at least $1 - \varepsilon$.

Because of conditional independence, we may choose the vertex sets $U^{(j)} = \pi_M(\mathcal{U}^{(j)})$ in the preceding discussion. Taking the intersection of all the events above, it follows that, with probability at least $1 - 2\varepsilon - o(1)$, there is a path in K_n between vertices 1 and 2 consisting of at most N_2 edges of FPP weight at most $f_n(X_{\max}^{(1)})$; a single edge of weight at most $f_n(X_{\max}^{(1)} + \varepsilon)$; at most $N_1 \log n$ edges of weight at most $g(\frac{1}{n}(1 + \varepsilon)) = f_n(1 + \varepsilon)$; a single edge of weight at most $f_n(X_{\max}^{(2)} + \varepsilon)$; and at most N_2 edges of weight at most $f_n(X_{\max}^{(2)})$. Therefore, with probability at least $1 - 2\varepsilon - o(1)$,

$$\begin{aligned} W_n &\leq (2N_2 + 2 + N_1 \log n) f_n((X_{\max}^{(1)} \vee X_{\max}^{(2)} \vee 1) + \varepsilon) \\ &\leq (2N_2 + 2 + N_1 \log n) f_n((M_0^{(1)} \vee M_0^{(2)}) + \varepsilon). \end{aligned} \quad (7.1)$$

To complete the proof, it suffices to show that the right-hand side of (7.1) is at most $f_n((M_0^{(1)} \vee M_0^{(2)}) + 2\varepsilon)$ with high probability. Since $M_0^{(1)}$ and $M_0^{(2)}$ do not depend on n and satisfy $M_0^{(j)} \geq 1$ a.s., it suffices to show that, for each fixed $x \geq 1$, we have $(2N_2 + 2 + N_1 \log n) f_n(x) \leq f_n(x + \varepsilon)$ for n sufficiently large. Taking $R = x + \varepsilon$ in [Condition 1.5 \(a\)](#), we can mimic the proof of [Lemma 3.1](#) to find that

$$\begin{aligned} \frac{f_n(x + \varepsilon)}{(2N_2 + 2 + N_1 \log n) f_n(x)} &\geq \frac{1}{(2N_2 + 2 + N_1 \log n)} \left(\frac{x + \varepsilon}{x} \right)^{\varepsilon_1 s_n} \\ &= \exp(\varepsilon_1 s_n \log(1 + \varepsilon/x) - \log \log n - O(1)). \end{aligned} \quad (7.2)$$

Since $s_n / \log \log n \rightarrow \infty$, the right-hand side of (7.2) diverges to infinity and therefore exceeds 1 for n sufficiently large. \square

Acknowledgements: A substantial part of this work has been done at Eurandom and Eindhoven University of Technology. ME and JG are grateful to both institutions for their hospitality. The work of JG was carried out in part while at Leiden University (supported in part by the European Research Council grant VARIS 267356), the Technion, and the University of Auckland (supported by the Marsden Fund, administered by the Royal Society of New Zealand) and JG thanks his hosts at those institutions for their hospitality. The work of RvdH is supported by the Netherlands Organisation for Scientific Research (NWO) through VICI grant 639.033.806 and the Gravitation NETWORKS grant 024.002.003.

A short guide to notation:

- $\text{SWT}_t^{(v)}$ is SWT from vertex v . We abbreviate $\text{SWT}^{(\varnothing_j)} = \text{SWT}^{(j)}$, $j \in \{1, 2\}$.
- $\mathcal{S}_t^{(j)}$ is the SWT from vertex $j \in \{1, 2\}$ such that $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ cannot merge and with an appropriate freezing procedure.
- $\mathcal{S}_t = \mathcal{S}_t^{(1)} \cup \mathcal{S}_t^{(2)}$
- $\text{BP}^{(j)}$ is branching process copy number j where $j \in \{1, 2\}$, without freezing.
- $\mathcal{B}^{(j)}$ is branching process copy number j where $j \in \{1, 2\}$, with freezing.
- \mathcal{B}_t is the union of 2 CTBPs with the appropriate freezing of one of them.

- $\tilde{\mathcal{B}}_t$ is the union of 2 CTBPs with the appropriate freezing of one of them, and the resulting thinning. Thus, $\tilde{\mathcal{B}}_t$ has the same law as the frozen \mathcal{S}_t .
- f_n is the function with $Y_e^{(\kappa_n)} \stackrel{d}{=} f_n(nE)$, where E is exponential with mean 1.
- μ_n is the image of the Lebesgue measure on $(0, \infty)$ under f_n
- λ_n is the exponential growth rate of the CTBP, cf. (2.13).

References

- [1] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont. The scaling limit of the minimum spanning tree of the complete graph. Preprint: arXiv:1301.1664v1 [math.PR], 2013.
- [2] Louigi Addario-Berry, Nicolas Broutin, and Bruce Reed. Critical random graphs and the structure of the minimal spanning tree. *Random Struct. Algor.*, 35(3):323–347, 2008.
- [3] Louigi Addario-Berry, Simon Griffiths, and Ross J. Kang. Invasion percolation on the Poisson-weighted infinite tree. *Ann. Appl. Probab.*, 22(3):931–970, 2012.
- [4] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.
- [5] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [6] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.
- [7] Omer Angel, Jesse Goodman, Frank den Hollander, and Gordon Slade. Invasion percolation on regular trees. *Ann. Probab.*, 36(2):420–466, 2008.
- [8] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6(23):1–13, 2001.
- [9] Shankar Bhamidi and Remco van der Hofstad. Weak disorder asymptotics in the stochastic mean-field model of distance. *Ann. Appl. Probab.*, 22(1):29–69, 2012.
- [10] Michael Damron and Artëm Sapozhnikov. Outlets of 2D invasion percolation and multiple-armed incipient infinite clusters. *Probab. Theory Rel. Fields*, 150(1-2):257–294, 2011.
- [11] Maren Eckhoff, Jesse Goodman, Remco van der Hofstad, and Francesca R. Nardi. Long paths in first passage percolation on the complete graph ii. branching process dynamics. Preprint: arXiv:1512.06145.
- [12] Maren Eckhoff, Jesse Goodman, Remco van der Hofstad, and Francesca R. Nardi. Short paths for first passage percolation on the complete graph. *J. Stat. Phys.*, 151:1056–1088, 2013.
- [13] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*, volume 33 of *Applications of Mathematics—Statistical Modelling and Applied Probability*. Springer-Verlag, Berlin, 1997.
- [14] Jesse Goodman. Exponential growth of ponds in invasion percolation on regular trees. *J. Stat. Phys.*, 147:919–941, 2012.

- [15] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.
- [16] Remco van der Hofstad. Random Graphs and Complex Networks. In preparation, see <http://www.win.tue.nl/~rhofstad/NotesRGCN.html>, 2014.
- [17] Daniel L. Stein and Charles M. Newman. Broken ergodicity and the geometry of rugged landscapes. *Phys. Rev. E*, 51(6):5228–5238, 1995.